



Trabajo Fin de Grado

Detección de Ataques de Epilepsia Mediante Sensores Inerciales

Grado en Ingeniería en Tecnologías de Telecomunicación

Autor:

Irene Tejado García

Tutor:

Victor Elvira Arregui

Título: Detección de Ataques de Epilepsia Mediante Sensores Inerciales

Autor: Irene Tejado García

Tutor: Victor Elvira Arregui

EL TRIBUNAL

Presidente: Francisco Javier González Serrano

Vocal: Pablo Basanta Val

Secretario: Francisco Javier Herraíz Martínez

Realizado el acto de defensa y lectura del Trabajo Fin de Grado el día 15 de Octubre de 2015 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de:

VOCAL

SECRETARIO

PRESIDENTE

Abstract

Epilepsy is the cause of a significant portion of the worlds disease burden. It often requires long-term care with a major medical expense but, due to the quick advancement in mobile technologies, there is an opportunity to acquire real-time data from patients.

This project aims to provide a solution to the automatic seizure detection and to implement a feasible medical device to monitor patients through the use of on-body wearable sensors, which have proved to be of great interest in this research area. For such purpose patients were admitted to long-term monitoring and the recorded data will be used for the automatic classification.

In this thesis we will focus on how the seizures can be detected using the different modules included in the sensor and the processing needed to prepare the data. A major emphasis will be devoted to the algorithms employed among which logistic regression, artificial neural networks and support vector machines are used.

Keywords: Epilepsy, sensor, monitoring, automatic classification.

Índice

Abstract	2
Lista de Figuras	7
Lista de Tablas	8
Acrónimos	9
1. Introduction	10
1.1. Goal	10
1.2. Motivation	10
1.3. Social and Economical Framework	10
1.4. Regulatory Framework	11
1.5. Structure of the Thesis	12
2. Estado del Arte	14
2.1. Definición de Epilepsia	14
2.2. Cuadro Clínico	14
2.3. Tratamiento	15
2.4. Monitorización	16
2.5. Detección Automática de Ataques de Epilepsia	18
3. Adquisición de Datos y Pre-procesado	20
3.1. Introducción	20
3.2. Sensor Shimmer	21
3.2.1. Unidades de Medición Inerciales	21
3.2.2. Obtención de la Señal Inercial	24
3.2.3. Calibración del Sensor	25
3.3. Recolección e Integración de Datos	26
3.4. Base de Datos	28
3.5. Calibración	29
3.6. Preprocesado	29
3.6.1. Eliminación de Outliers	30
3.6.2. Filtro Paso Bajo	31
3.7. Extracción y Selección de Características	32

4. Procesado de las Señales	35
4.1. Introducción	35
4.2. Algoritmos Propuestos	36
4.2.1. Aproximación Mediante Umbral	36
4.2.2. Regresión Logística Binaria	38
4.2.3. Redes Neuronales	41
4.2.4. Máquinas de Vectores de Soporte	47
4.3. Medidas de rendimiento	54
5. Resultados	58
5.1. Regresión Logística Binaria	58
5.2. Redes Neuronales	61
5.3. Máquinas de Vectores de Soporte	65
5.4. Resumen y Comparativa de los Modelos Propuestos . .	67
6. Conclusions	69
6.1. General conclusions	69
6.2. Future Work	70
A. Planificación y Presupuesto	71
A.1. Planificación del Proyecto	71
A.2. Diagrama de Gantt	71
A.3. Presupuesto	73
A.3.1. Costes Materiales	73
A.3.2. Costes de Personal	73
A.3.3. Costes Totales	73
B. Summary: Epileptic Seizure Detection Through Inertial Sensors	75
B.1. Introduction	75
B.1.1. Shimmer Sensor	75
B.1.2. Data Collection	76
B.1.3. Data Base	77
B.1.4. Calibration	77
B.1.5. Preprocessing	78
B.1.6. Features Extraction	79
B.2. Signal Processing	79
B.2.1. Logistic Regression	80
B.2.2. Neural Networks	80

B.2.3. Support Vector Machines	81
B.2.4. Performance Measures	82
B.3. Results	83
B.4. Conclusions and Future Work	84

Índice de figuras

1.	Nervio Vago.	16
2.	Grabación vídeo-encefalográfica.	17
3.	Diagrama con la metodología de trabajo	20
4.	Sistema de Coordenadas del sensor Shimmer.	21
5.	Giróscopo uni-axial sobre mesa rotatoria	23
6.	Base de calibración del sensor Shimmer	25
7.	parámetros de calibración	26
8.	Maletín de Monitorización	26
9.	Elementos del maletín	27
10.	Base del Sensor	27
11.	Calibración del sensor	30
12.	Presencia de outliers	31
13.	Acelerómetro después de la eliminación de Outliers	31
14.	Filtrado	32
15.	Aplicación de un umbral al acelerómetro	37
16.	Detección de ataques a través de un umbral aplicado al acelerómetro.	38
17.	Función logistica con $\beta_0 + \beta_1 x$ en el eje horizontal y p_i en el eje vertical.	39
18.	Neurona.	41
19.	Modelo de Red Neuronal con una capa y múltiples neuronas.	42
20.	Red Neuronal con múltiples capas.	43
21.	Interfaz de MATLAB.	44
22.	Interfaz de MATLAB.	44
23.	Interfaz de MATLAB.	45
24.	Interfaz de MATLAB.	46
25.	Interfaz de MATLAB.	46
26.	Hiperplano de las SVMs.	47
27.	Conjunto e hiperplano separador. Los puntos huecos representan los vectores con etiqueta $y = 1$ y los rellenos $y = -1$	49
28.	Matriz de confusión	55
29.	Curvas ROC	56
30.	Ejemplo de detección con Regresión Logística.	61

31.	Matriz de confusión para una Red Neuronal con ventana deslizante de 20 muestras.	63
32.	Matriz de confusión para una Red Neuronal con ventana deslizante de 100 muestras.	64
A.1.	Diagrama GANTT del proyecto.	72
B.1.	Methodology Flowchart	76
B.2.	Recorder Case	77
B.3.	Logistic Function	80
B.4.	Multiple layer Neural Network.	81
B.5.	Set of data and hyperplane	82
B.6.	ROC curve	83

Índice de Tablas

1.	Base de Datos	28
2.	Clasificación de las muestras de test utilizando Regresión Logística: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.	59
3.	Clasificación de las muestras de test utilizando Regresión Logística: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.	60
4.	Clasificación de las muestras de test utilizando SVM: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.	65
5.	Clasificación de las muestras de test utilizando SVM: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.	66
6.	Resumen de Resultados	67
7.	Planificación del proyecto.	71
8.	Costes Materiales	73
9.	Costes de Personal	73
10.	Costes Totales	74
11.	Data Base	77
12.	Results Summary	83

Lista de Acrónimos

ACC	Acelerómetro
GYR	Giróscopo
MAG	Magnetómetro
EEG	Electroencefalograma
ECG	Electrocardiograma
EMG	Electromiograma
IMU	Unidades de Medición Inerciales
RL	Regresión Logística
SVM	Máquinas de Vectores de Soporte
RN	Red Neuronal
ROC	Curva Operativa del Receptor
ERM	Minimizacon del Riesgo Empírico
SRM	Minimización del Riesgo Estructural

1. Introduction

This chapter aims to introduce the framework upon which this project is developed. There will be relevant aspects regarding the motivation and goal for the realization of this thesis. Later on, the general structure of this paper will be presented so as to help the reader localize the topics that will be explained.

1.1. Goal

The objective of this thesis is the development of an algorithm capable of detecting epileptic seizures based on the data extracted from a wearable sensor placed on the patients under the experiment.

1.2. Motivation

Epilepsy is cataloged as one of the most common neurological disorders and occurs with an incidence of **68.8/100,000 person-years** at risk [1]. Moreover, the number of patients with this disease is increasing over the years. According to a European Brain Council report, brain disorders, including epilepsy, cost Europe almost **€800 billion a year**. In this kind of diseases, patients often require long-term care. Systems employed to automatically detect certain physical activity are highly attractive in terms of healthcare monitoring since despite new treatments, the number of seizures is not fully controlled.

The information obtained from this kind of studies is valuable for long-term assessment of physiological diseases and the realization of accurate seizure diaries providing customized therapies based on the frequency of the attacks.

In this thesis, it will be discusses all the topics related to the approach of automatic seizures detection: sensor selection, data acquisition, features selection and extraction, and finally classification.

1.3. Social and Economical Framework

As one of the most common neurological conditions, epilepsy has been highlighted as a priority by the **Organización Mundial de la**

Salud (OMS). Many brain disorders are chronic and incurable diseases whose effects can continue for years. Most of the deaths caused by epilepsy are potentially avoidable and there are patients who could achieve seizure freedom with the appropriate treatment.

Recent research shows an average waiting time of weeks for a specialist appointment or diagnosis. With these long periods, patients are being put at risk. Some of the demands for the government by national institutes and organizations include the increase of personal or the creation of a clearer customized care.

From the technological point of view, the quick advancement in mobile technologies has brought the opportunity to acquire process real-time data from many patients spending large periods of time hospitalized, in order to find the way to create a precise and efficient method to monitoring those patients from home to improve their life quality.

The economical cost of such an expanded disease is huge so studies in this area are particularly important in order to provide an idea of how to minimize it. On the one hand, direct costs are mainly related to health services such as drugs, diagnostic tests and consultations. On the other hand, indirect costs derived from the change in the productive capacity of the individual work due to the disease. Nevertheless, the main economic consequence for most patients is to suffer the limitation in their work activity if seizures are not controlled.

1.4. Regulatory Framework

When patients go to the doctor for health assistance, their data is usually used for more purposes than just the diagnosis. For this reason is necessary to understand the existence of an agreement where patients decide what the owner of the clinic can do with their data. **La Ley Orgánica de Protección de Datos** [2] is created in order to ensure that protection.

However, this law does not include a definition of health data. To solve this issue, it is considered that personal data relating to health includes both physical and mental information.

The Organic Law of Data Protection establishes a set of mandatory principles that those responsible for the files should meet:

- The purpose of the study and data collection should be determined and reported to the patient. Data should be updated and when no longer necessary, deleted.
- Patients should be informed of the existence of a file, its purpose and the recipients of the information as well as their address and identity in case they need to exercise their rights.
- In accordance with the **Real Decreto 994/1999 por la Regulación de Medidas de Seguridad**, files must fulfill the security measures and a security document must also be available.
- The owner of the clinic and its employees are required to keep professional secrecy.
- Whenever the data is going to be used by a third party, the person concerned must be informed and asked for their consent.

1.5. Structure of the Thesis

This document is divided in 6 chapters in which all the necessary and related issues for the development of this thesis are explained.

In this Chapter, a brief **introduction** about the project is done. The goal, motivation and frameworks are commented as well.

Chapter 2 will cover everything related to **epilepsy**. From its definition as a disease, the clinical symptoms, the diagnosis and the current treatments as well as other alternatives for those patients not responding to regular treatments. Afterwards, it will be explained the necessary monitoring of the patients and how automatic detection algorithms are being created for such purpose.

In Chapter 3, it is explained how the **acquisition of the data** was managed as well as a discussion about the sensor choice and its functional characteristics. Later on, there is an explanation about the first steps performed on the data during the **preprocessing** phase.

Finally, the selection and extraction of the relevant characteristics will be discussed.

Chapter 4 focuses on the **signal processing** and all the theoretical concepts necessary for understanding the performance of the proposed models for the realization of an algorithm capable of automatically detect epileptic seizures. Moreover, it is also explained some concepts regarding the performance measures of those models

Finally, in Chapters 5 and 6 there will be a discussion about the **results** obtained and general **conclusions** about the thesis including **future work** will be mentioned.

2. Estado del Arte

En este capítulo se hace una introducción a la epilepsia, tanto a su definición como a los diferentes tipos que existen. Se presentan las características de sus síntomas así como las diferentes maneras de diagnóstico y tratamiento. Se explica, además, la necesidad de distintos métodos de monitorización para pacientes con epilepsia. Finalmente, se comentan algunos métodos de detección automática que se han llevado a cabo.

2.1. Definición de Epilepsia

La epilepsia [3] es un trastorno neurológico basado en la disfunción del sistema nervioso central y es provocado por un desequilibrio en la actividad eléctrica de las neuronas. Afecta a casi el 1 % de la población mundial siendo uno de los desórdenes patológicos con más alto índice de prevalencia. La manifestación de la epilepsia se denomina síndrome epiléptico y comprende un espectro muy amplio de cuadros epilépticos así como de tratamientos.

2.2. Cuadro Clínico

Una crisis epiléptica se caracteriza por una actividad neuronal anormal que altera el correcto funcionamiento del cerebro y provoca descargas eléctricas sincrónicas. El tipo de síntoma manifestado dependerá de la parte del cerebro afectada. Estas crisis duran desde segundos hasta varios minutos y pueden incluso llegar a propagarse de unas zonas del cerebro a otras. Si nos basamos en la zona del cerebro afectada, se pueden distinguir las siguientes categorías de crisis epilépticas:

- **Ausencia:** la consciencia se ve alterada y la persona permanece inmóvil con el conocimiento perdido durante unos segundos.
- **Clónica:** se caracteriza por producir contracciones musculares rítmicas y repetitivas durante la cual se puede producir mordedura de la lengua o lesiones a causa de una caída.
- **Tónico-Clónica:** comienza por estiramiento espontáneo de los músculos, seguida por una segunda fase clónica lo que provoca una caída al suelo acompañada de movimientos rítmicos de todo

el cuerpo. Este tipo de crisis son, sin duda, las más aparatosas aunque si se toma la medicación adecuada son fácilmente controlables.

En muchos casos, existen personas que padecen epilepsia pero no llegan a mostrar síntomas por lo que es importante diferenciar entre epilepsia y síndrome epiléptico. Cada tipo de epilepsia tiene diferentes manifestaciones. Existen ataques tan leves que son apenas percibidos. Para el desarrollo de este trabajo fin de grado, tendremos en cuenta aquellas crisis que presentan movimientos motores o convulsiones, ya que serán aquellas que el sensor pueda registrar.

Las *crisis hipermotoras* [4] son aquellas que se manifiestan de manera violenta con movimientos incontrolados de brazos y piernas, que pueden durar desde unos segundos a varios minutos. Estos ataques se atribuyen a crisis originadas en el lóbulo frontal y suelen aparecer durante el sueño. Después de la crisis, el paciente puede sufrir confusión durante un breve periodo de tiempo. Debido a la violencia con la que ocurren las crisis hipermotoras, el riesgo para el paciente es muy alto por lo que se hace necesario detectar todas las crisis de este tipo.

2.3. Tratamiento

Tanto la epilepsia como las crisis epilépticas tienen en la actualidad tratamiento y en casi el 70 % de los pacientes, la medicación da resultado. Sin embargo, cuando el tratamiento no funciona, el enfoque más utilizado es el uso de medicamentos antiepilépticos, en su mayoría con efecto sedante. La toma de varios tipos de medicamentos a veces conlleva a padecer efectos secundarios y disminuye la probabilidad de éxito del fármaco.

Existen algunos tipos de epilepsia resistentes a los tratamientos. Se denominan *epilepsias refractarias* [5] y son un gran problema clínico. Los pacientes que sufren este tipo de desorden son tratados con una combinación de diferentes medicamentos. Para estos casos, el neurólogo es el encargado de determinar si el paciente necesita cirugía. Hay varias alternativas que se están llevando a cabo, aún en fases de pruebas, como la implantación de electrodos en el cerebro capaces de enviar

señales eléctricas a la parte del cerebro responsable de las crisis con el objetivo de lograr su simulación.

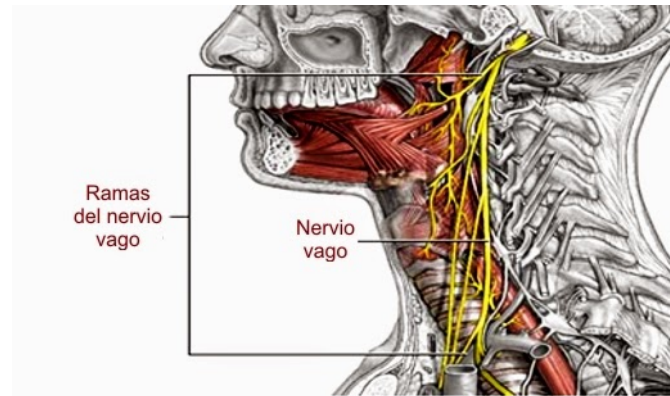


Figura 1: Nervio Vago.

Otra posible solución que se puede aplicar para los casos de epilepsia refractaria es la *estimulación eléctrica del nervio vago* [6]. Se utiliza sobre pacientes que no son candidatos a ser quirúrgicamente intervenidos o que, aún siendolo, no desean someterse a la operación. La forma de llevar a cabo este procedimiento consiste en la inserción de un electrodo sobre el nervio vago ya que posee muchas conexiones con distintas partes del cerebro.

2.4. Monitorización

Una de las cuestiones más importantes que surgen al lidiar con un paciente de epilepsia es cómo va a tratarse a largo plazo su asistencia médica. Diversos estudios [7] mencionan la necesidad de mejorar el sistema para poder proporcionar mejores cuidados a los pacientes, quienes a su vez, deben estar debidamente educados sobre su enfermedad.

El objetivo de todo paciente es el de ser capaz de controlar las crisis a través de un tratamiento eficaz y poder mejorar su calidad de vida. De esta forma se deduce que cualquier seguimiento de un paciente con epilepsia debe centrarse en dos objetivos: el **control de las crisis** y el **control de los efectos adversos** que se puedan sufrir a causa de la medicación. Durante las revisiones, el neurólogo debe encargarse de

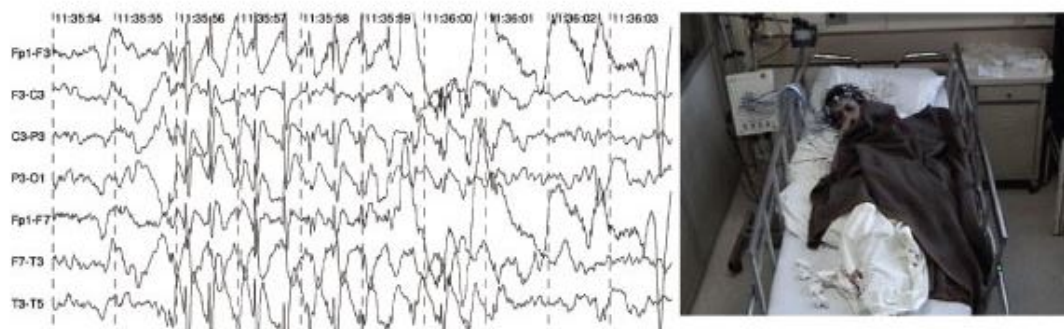


Figura 2: Grabación vídeo-encefalográfica.

comprobar el grado de cumplimiento terapéutico así como el posible consumo de tóxicos por parte del paciente.

Actualmente, ya sea debido al grado de epilepsia o a la irregularidad con la que sufran las crisis, hay muchos pacientes que pasan largas temporadas hospitalizados para la realización de una monitorización prolongada. Las características de las convulsiones y su frecuencia pueden ser evaluadas mediante la apropiada monitorización. El método más extendido es el examen de las ondas cerebrales a través de electroencefalograma (EEG), mediante el cual se mide la actividad eléctrica del cerebro [8]. Para su realización, se han de colocar unos electrodos a lo largo de todo el cuero cabelludo del paciente. Los electrodos están conectados a una grabadora que convierte las señales eléctricas en patrones para facilitar su posterior interpretación. Sin embargo, esta técnica requiere personal especializado, no pudiendo dejar al paciente sin atención. Además, el EEG se caracteriza por tener un nivel bajo de sensibilidad.

El uso de EEG puede combinarse con la **monitorización por vídeo** [9]. De esta forma se consigue un registro simultáneo del estado del paciente y de su actividad cerebral. El desarrollo de esta técnica se ha expandido sobre todo en los últimos años y debido al avance tecnológico y al gran abaratamiento de los costes, es una técnica ampliamente usada.

Para evitar que los pacientes pasen largos periodos de tiempo hospitalizados, existen algunos métodos que permiten reproducir en casa un ambiente de monitorización a distancia. En estos casos, los senso-

res son los encargados de realizar la tarea de seguimiento del paciente a través de, por ejemplo, acelerómetros. Uno de los principales retos en cuanto al uso de sensores que permitan la **monitorización desde casa** [10] es la aceptación de esa tecnología por parte del paciente. Es importante encontrar un equilibrio entre la relevancia al estudio que aporte la tecnología y la comodidad para el paciente.

Este método de monitorización ha supuesto una gran reducción de la tasa de mortalidad así como de pacientes que llegaban de urgencia al hospital. Esto le añade un valor extra al cuidado médico además de disminuir los gastos hospitalarios.

Este trabajo se centra en la aplicación de sensores inerciales para desarrollar algoritmos de detección de ataques de epilepsia. Especialmente útil para que aquellas personas que sufren epilepsia crónica, afronten mejor su enfermedad y sea posible llevar a cabo un seguimiento más exhaustivo sobre su evolución.

2.5. Detección Automática de Ataques de Epilepsia

Para poder realizar un diagnóstico, no es necesario disponer de información sobre la cantidad de ataques que ha sufrido un paciente ni a que hora sucedieron. Sin embargo, si se quiere desarrollar un algoritmo capaz de detectar y alarmar automáticamente de estos ataques, esa información se vuelve necesaria.

Durante este trabajo, nos centraremos en la detección de ataques de epilepsia a través de sensores inerciales que miden la aceleración (ACC), la orientación y la señal magnética. Sin embargo, existen otros métodos que también se utilizan y se introducirán en este apartado.

Algunos estudios [11] presentan el **electromiograma** (EMG) como un posible método de detección. El EMG mide la tensión muscular por lo que da una amplia información sobre la actividad motora. Especialmente útil para las crisis tónicas ya que los músculos son los principales afectados.

Otros métodos no tienen en cuenta el movimiento y se basan en técnicas como las **señales auditivas** para la detección. Los gritos, risas, etc... son comunes en los pacientes cuando están sufriendo una crisis.

El **electrocardiograma** (ECG) es muy utilizado en la medicina y mide la actividad eléctrica del corazón. A través de los latidos y de las variaciones que estos sufran se puede identificar una crisis. Para ello, el paciente debe colocarse unos electrodos en el pecho.

El uso de **acelerómetros**, como el utilizado para el desarrollo de este trabajo, es un método muy extendido para el reconocimiento y detección de actividad humana. Los métodos de clasificación usados con esta técnica son, por ejemplo, los arboles de decisión, clustering [12], los clasificadores binarios y las cadenas de Markov, entre muchos otros. Durante la última década, la investigación acerca de la detección de crisis de epilepsia contempla el uso de acelerómetros y el vídeo como complemento, siendo esta una manera menos intrusiva para el paciente. Este es el procedimiento que se ha seguido en este trabajo y más detalles sobre el sensor utilizado se explican en el Apartado 3.2, así como de los algoritmos implementados para la detección automática de ataques en el Apartado 4.

3. Adquisición de Datos y Pre-procesado

Este Capítulo se divide en dos partes: en la primera, se trata todo lo relacionado con la adquisición de los datos y se hace una introducción acerca de su naturaleza para la realización de algoritmos automáticos de detección y del sensor escogido. Se presenta la necesidad de crear una base de datos para organizar de manera eficiente a los pacientes. En la segunda parte del Capítulo se verá el pre-procesado llevado a cabo sobre las señales obtenidas del sensor así como la extracción de características.

3.1. Introducción

Los datos disponibles para la realización de este trabajo han sido facilitados por la Unidad de Epilepsia de la Fundación Jimenez Díaz, a cargo del doctor José Serrasota. Se cuenta con grabaciones realizadas durante las horas de sueño de los pacientes que a su vez, llevaban los sensores colocados en las muñecas.

El detector de ataques incorpora una secuencia de pasos que, aplicados sobre los datos, facilitan el desarrollo del sistema y mejoran sus resultados. El método de trabajo seguido se divide en varias partes, que se llevarán a cabo tanto para el conjunto de entrenamiento como para el de test.

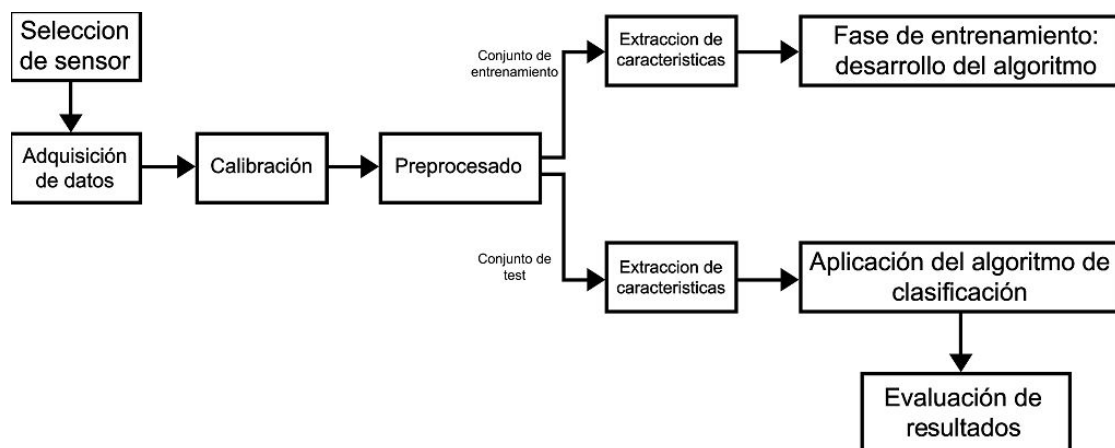


Figura 3: Diagrama con la metodología de trabajo

Uno de los aspectos más importantes a la hora de elaborar un sistema de detección automática, es la elección del sensor. Para mayor

comodidad del paciente, el sensor debe ser pequeño y ligero de manera que no comprometa el desarrollo normal del estudio. Ya que los datos se recolectaron durante horas de sueño, es importante que esa situación se mantuviese estable y hubiese las mínimas interrupciones posibles. La visualización del paciente debe ser clara. De no ser así, las marcas sobre la grabación podrían no realizarse correctamente.

3.2. Sensor Shimmer

El sensor portátil Shimmer [13] permite la captura, transmisión y procesamiento de datos corporales detectados en tiempo real. Contiene el equipamiento y las herramientas necesarias para llevar a cabo todas las etapas del proceso permitiendo así el desarrollo de una solución completa del sensor.

3.2.1. Unidades de Medición Inerciales

El sensor inalámbrico Shimmer contiene un dispositivo electrónico denominado IMU, que mide la velocidad, orientación y fuerzas gravitacionales. Para ello utiliza un acelerómetro (ACC), un giróscopo (GYR) y un magnetómetro (MAG). De cada uno de los módulos se obtienen lecturas sobre los ejes inerciales X,Y y Z. Sin embargo, la orientación del sistema de coordenadas varía en función del módulo, tal como se ilustra en la Figura 9.

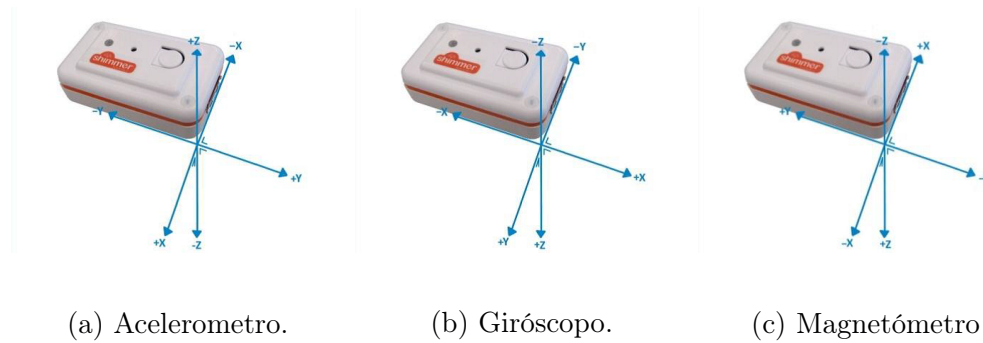


Figura 4: Sistema de Coordenadas del sensor Shimmer.

Acelerómetro

Un ACC es un dispositivo que mide la aceleración de un cuerpo debido a todas las fuerzas que actúan sobre él. Su unidad en el Sistema

Internacional es m/s^2 . La aceleración tiene dos componentes principales: la aceleración inercial y gravitacional.

■ Aceleración Inercial

La aceleración inercial se debe a la aplicación de cualquier fuerza, exceptuando la de la gravedad, sobre un cuerpo. Cuando un cuerpo este está quieto o se mueve con velocidad constante consideraremos que hay fuerzas inerciales actuando sobre él.

■ Aceleración Gravitacional

La gravedad es un concepto que se refiere a la alteración de la velocidad de un cuerpo debido a la acción sobre él de la fuerza de la gravedad. Es un fenómeno por el cuál dos cuerpos se atraen con una fuerza propocional a sus masas.

Cuando un acelerometro esta completamente estacionario, este mide una aceleración constante igual en magnitud que la aceleración de la gravedad. A menudo nos referiremos a esa magnitud como g .

Acelerometro Tri-axial

El sensor SHIMMER cuenta con un ACC triaxial formado por tres ACC uni-axiales mutuamente ortogonales. Cada uno de los ejes del ACC nos aporta información sobre la aceleración a la que está siendo sometido el sensor en su respectivo eje. El vector de la aceleración depende de los ángulos formados entre los ejes y la direccion de los componentes de la aceleración:

$$\underline{a} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} = \begin{pmatrix} a_I \cos(\theta_x) + g \cos(\phi_x) \\ a_I \cos(\theta_y) + g \cos(\phi_y) \\ a_I \cos(\theta_z) + g \cos(\phi_z) \end{pmatrix}$$

Gir6scopo

Un GYR es un dispositivo que se usa para medir la velocidad de rotaci6n que sufre un objeto en el eje o ejes sobre el que esta rotando.

Gir6scopo Tri-axial

Un GYR tri-axial, como el que contiene el sensor SHIMMER, est1 formado por tres GYRs uni-axiales colocados de manera ortogonal, como ya ocurr1a con el aceler6metro, que permiten obtener un vector \underline{w} de la velocidad angular:

$$\underline{w} = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} = \begin{pmatrix} w \cos(\beta_x) \\ w \cos(\beta_y) \\ w \cos(\beta_z) \end{pmatrix}$$

donde $\beta_x, \beta_y, \beta_z$ son los 1ngulos formados entre los ejes de medida y la componente de la velocidad angular en cada eje respectivamente.

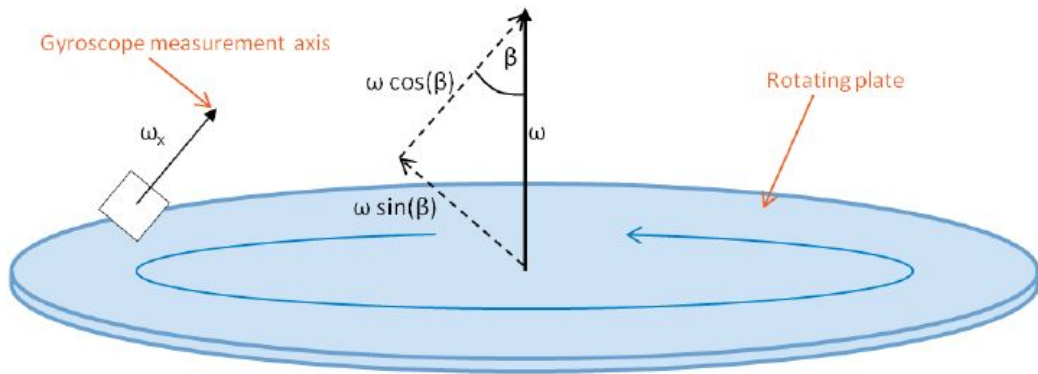


Figura 5: Gir6scopo uni-axial sobre mesa rotatoria

Magnet6metro

Un *campo magn6tico* es la descripci6n matem1tica de la influencia que tienen las corrientes y materiales magn6ticos y puede darnos informaci6n sobre la direcci6n y conducta de un objeto.

Para medir la direcci6n y fuerza de un campo magn6tico se hace uso de un MAG.

Magnetómetro Tri-axial

Un MAG tri-axial está formado por tres MAGs uni-axiales colocados de forma ortogonal. El vector del campo magnético se obtiene de la siguiente manera:

$$\underline{m} = \begin{pmatrix} m_x \\ m_y \\ m_z \end{pmatrix} = \begin{pmatrix} m \cos(\alpha_x) \\ m \cos(\alpha_y) \\ m \cos(\alpha_z) \end{pmatrix}$$

donde $\alpha_x, \alpha_y, \alpha_z$ son los ángulos formados entre los ejes de medida y la componentes del campo magnético en cada eje respectivamente.

3.2.2. Obtención de la Señal Inercial

Las unidades de medición inerciales (IMU), como en todos los sensores, se ven perjudicadas por errores tales como el ruido, error de sesgo o errores de alineamiento con los ejes. Además, las medidas capturadas por el sensor están limitadas por valores dentro de un rango finito mientras que los fenómenos físicos que se observan no lo están. En consecuencia, los valores medidos por cualquier sensor no son exactamente iguales a los valores del fenómeno medido.

Medición de un Sensor Tri-axial

Un sensor tri-axial debería consistir, de manera ideal, en el conjunto de tres sensores uni-axiales ortogonales entre si. Sin embargo, los sensores que se suelen utilizar tienden a sufrir errores debido a una mala alineación en su posicionamiento por lo que los ejes del sensor dejan de ser mutuamente ortogonales.

Por estos motivos, es necesario definir una matriz de rotación que relacione los valores tomados de la sensibilidad en los distintos ejes inerciales con los valores reales. El cálculo de dicha matriz se realiza durante el proceso de calibración y se define con la matriz R .

Para un sensor tri-axial y el vector del fenómeno a detectar \underline{v} , la captura final de datos se formula de la siguiente manera:

$$\underline{Y} = KR\underline{v} + \underline{b} + \underline{n} \quad (1)$$

donde \underline{Y} es el vector salida del sensor, \underline{v} es el vector del fenómeno

a detectar, K es la matriz diagonal con los valores de escalado de los ejes, R es la matriz rotacional que define los ejes reales (x,y,z), \underline{b} es el vector del margen del sesgo y \underline{n} es el vector del ruido.

3.2.3. Calibración del Sensor

El sensor SHIMMER cuenta con una aplicación que hace posible la obtención de los parámetros para la CAL. De la Fórmula 1 ya mencionada, la matriz K , R y b se obtendrán a partir de dicha aplicación. El sensor se coloca sobre una base de calibración como la mostrada en la Figura 6 y se conecta al ordenador.



Figura 6: Base de calibración del sensor Shimmer

En el panel principal de la aplicación se puede encontrar información acerca de los parámetros de calibración que presenta el sensor conectado y los que se encuentran almacenados en la memoria del sensor. Además, se dispone de una serie de botones para controlar esos parámetros. Para cada uno de los módulos del sensor, se deben obtener los parámetros de calibración e introducirlos en la memoria del sensor para que el resultado que se obtenga de la captura pueda ser procesado correctamente. En la Figura 7 se muestra un ejemplo de los parámetros obtenidos para el ACC. Estos parámetros son almacenados en un fichero que se usará previamente al procesamiento de la señal obtenida del sensor para su calibración.

Accelerometer Calibration Parameters Stored in Shimmer Memory								
Offset Vector (ba)			Sensitivity Matrix (Ka)			Alignment Matrix (Ra)		
-1			65535	0	0	-0.01	-0.01	-0.01
-1			0	65535	0	-0.01	-0.01	-0.01
-1			0	0	65535	-0.01	-0.01	-0.01

Figura 7: parámetros de calibración

3.3. Recolección e Integración de Datos

Para la recolección de datos se ha diseñado un prototipo de monitorización del que se pueden diferenciar dos partes: por un lado el maletín registrador, el cual contiene todo lo necesario para la toma de datos, tal como ilustra la Figura 8. Por otro lado, se dispone de un ordenador cuya función es ejecutar la aplicación encargada de realizar y mostrar el procesamiento de la información que se ha grabado en el maletín.



Figura 8: Maletín de Monitorización

- **Pantalla:** el maletín contiene un ordenador con Sistema Operativo Windows 7 donde se pueden visualizar los datos del paciente, del sensor o de la grabación.
- A través del **teclado inalámbrico** se interactúa con el ordenador.

- **Botones:** se dispone de tres botones con los que se puede tanto activar como desactivar el equipo, la grabación y la pantalla.
- **Cámara de vídeo:** permite la grabación de los pacientes. Los vídeos se almacenan en el ordenador del maletín.



(a) Ordenador con pantalla incorporada.



(b) Cámara.

Figura 9: Elementos del maletín

- **Base del Sensor:** permite la carga de los sensores así como el registro de los datos tomados.



Figura 10: Base del Sensor

Para lograr el correcto funcionamiento del prototipo, las enfermeras de la Fundación Jiménez Díaz fueron entrenadas siguiendo un tutorial de utilización proporcionado por la Universidad. Dicho tutorial incluye los pasos necesarios para la conexión de todos los aparatos del maletín y una vez conseguido, explica el proceso de recogida de datos.

El acceso a la información grabada se hace a través de la aplicación ya mencionada para importar al ordenador los ficheros generados por el maletín. Para el desarrollo de este proyecto se dispone de 3 fuentes de información:

Fuente	Formato de Almacenamiento
Medidas de movimiento	Historia_Sesion_fecha_Hora.A01 Historia_Sesion_fecha_Hora.A02
Vídeo y Audio	Historia_Sesion_fecha_Hora.AVI
Marcas temporales	Historia_Sesion_fecha_Hora.BMK

3.4. Base de Datos

La organización de los pacientes se realiza a través de la implementación de una Base de Datos. Debido a la Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal, no se pueden exponer los datos de los pacientes por lo que se les asigna un número identificativo y un número de registro por cada grabación que poseen.

ID	Reg	Nombre del fichero	A01	A02	Fecha	Hora Inicio Grabación	Hora crisis Monitor
2	1	ID2_R1_14-054_ 20140223_230213-634	0	1	23-02-14	23:02:13	
3	1	ID3_R1_14-001_ 20140127_173033-058	1	0	27-01-14	17:30:33	
...	...						
10	2	ID10_R2_00-000_ 20130920_050612-654	1	1	20-09-13	5:06:12	5:38:40 6:02:40 7:04:55 7:59:35 8:25:09

Tabla 1: Base de Datos

De cada paciente se han almacenado los siguientes datos:

- **ID:** número de identificación del paciente.
- **Registro:** registro de grabación del paciente.
- **Nombre del fichero:** A cada fichero se le asigna un nombre en función del ID, registro del paciente, fecha y hora de grabación.

- **A01, A02:** Son los ficheros de grabación recolectados a través del sensor. Como no todos los pacientes poseen ambos ficheros, se asigna un 0 en la casilla correspondiente si no se dispone del fichero, o por el contrario, un 1 si se dispone de él.
- **Fecha:** fecha en la que se tomaron los datos del paciente.
- **Hora de inicio:** Se almacena la hora de inicio de grabación para poder situar en el tiempo las posibles crisis que sufra el paciente de cara al procesado y desarrollo del algoritmo de detección.
- **Hora crisis:** De haberse producido una crisis, se apunta la hora aproximada en la que sucedió.

En la Tabla 1 se observa la información almacenada de algunos pacientes en la base de datos.

3.5. Calibración

Una vez se ha organizado la información de los pacientes, el primer paso realizado sobre los datos en crudo es el de calibración. Este proceso se lleva a cabo tal como se explica en el Apartado 3.2.2 y se aplica sobre todos los componentes del sensor ya que pueden verse afectados por diferentes tipos de error.

Cuando se adquieren los datos del sensor, estos se obtienen en unas unidades de medida que no son las apropiadas para su procesado. A través de la calibración conseguimos obtener valores estándar (m/s^2 , $grados/seg$) y unos datos calibrados.

En la parte superior de la Figura 11 se ilustra un ejemplo del ACC sin calibrar, mientras que en la parte inferior los datos se encuentran ya calibrados. La diferencia fundamental son las unidades de medida. En los datos calibrados obtenemos valores estándar y manejables para el procesado. La forma de la señal se mantiene para que el estudio de los datos sea correcto.

3.6. Preprocesado

El propósito principal del preprocesado es manipular y transformar la información en bruto en un conjunto de datos más fácilmente accesible. Esta etapa engloba todas aquellas técnicas de análisis que permiten mejorar la calidad de los datos. Se encarga de segmentar y

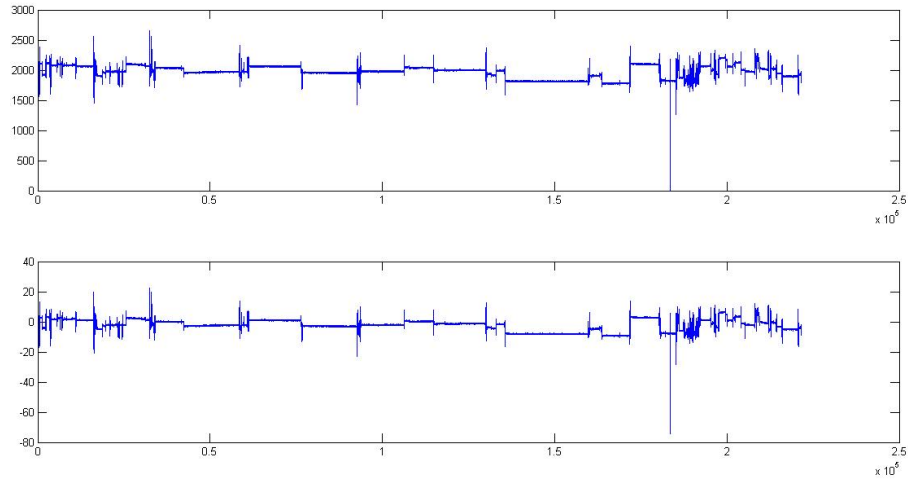


Figura 11: Calibración del sensor

eliminar información redundante.

El preprocesamiento permite aplicar los modelos de aprendizaje de forma más rápida y sencilla, obteniendo una mayor precisión. Los pasos que se van a seguir durante esta etapa incluyen la resolución de conflictos entre datos como, por ejemplo, la eliminación de outliers y la aplicación de un filtro paso bajo.

3.6.1. Eliminación de Outliers

Los outliers se definen como observaciones que numéricamente distan mucho del resto de datos. Generalmente estos valores se relacionan con errores o información engañosa y se deben a múltiples causas anómalas. Por esta razón, son excluidos del conjunto de datos.

A través de las señales obtenidas de los sensores se deduce que los outliers no son mas que picos de duración mínima debidos a una malfunción transitoria del sensor. En comparación con los picos que representan ataques epilépticos, los outliers alcanzan valores mucho más altos y si no se eliminasen podrían llevar a una clasificación errónea de los ataques. Aplicando un umbral, se identifican estos picos tan desproporcionados y se sustituyen por un valor medio de la señal en las posiciones contiguas al outlier. En la Figura 12 se puede identificar la presencia de un outlier sobre los datos obtenidos del ACC y se puede

ver la diferencia de altura en comparación con el resto de picos que podrían ser identificados como ataques.

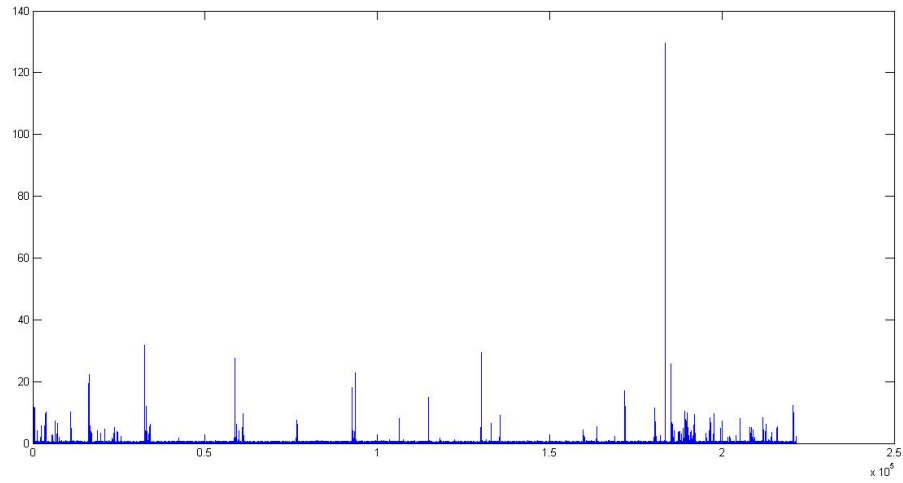


Figura 12: Presencia de outliers

Después de eliminar los outliers el ACC queda tal como se ilustra en la Figura 13. Los picos restantes tienen un nivel similar y son útiles para el resto del procesado y la detección de ataques.

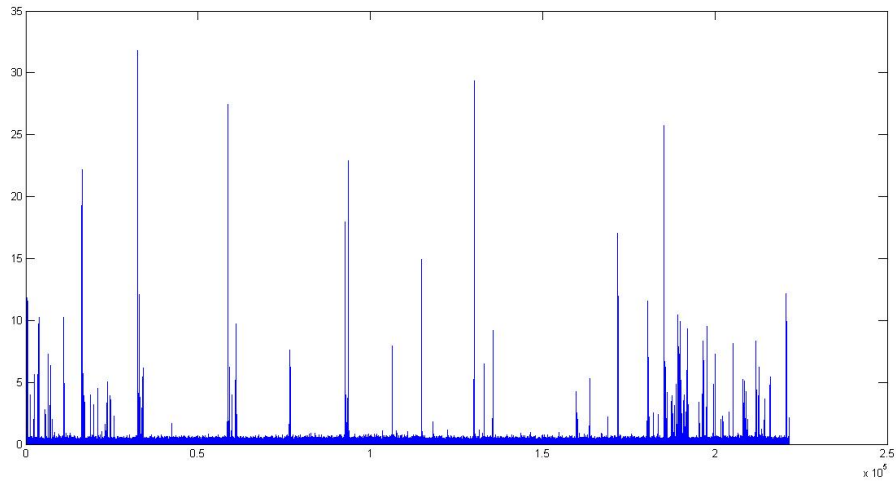


Figura 13: Acelerómetro después de la eliminación de Outliers

3.6.2. Filtro Paso Bajo

Las señales que se obtienen hasta el momento contienen picos de energía muy débiles que no aportan información de cara a extraer características para la clasificación de patrones. El siguiente paso que se

realiza tiene como objetivo eliminar estos picos y para ello, se ha decidido utilizar un filtro paso bajo

Un filtro paso bajo [14] se caracteriza por permitir el paso de frecuencias bajas y atenuar las frecuencias altas a través de la definición de una frecuencia de corte.

Para el preprocesado se ha aplicado un filtro de Butterworth. Este tipo de filtros se caracterizan por mantener la señal constante casi hasta la frecuencia de corte, que en este caso es:

$$f_c = 0,01Hz \quad (2)$$

Todos los componentes de la señal cuya frecuencia esté por encima de la de corte, serán atenuados. Además, se consigue que el nivel del ruido quede eliminado. En la Figura 14 podemos ver el resultado de aplicar el filtro paso bajo sobre el ACC.

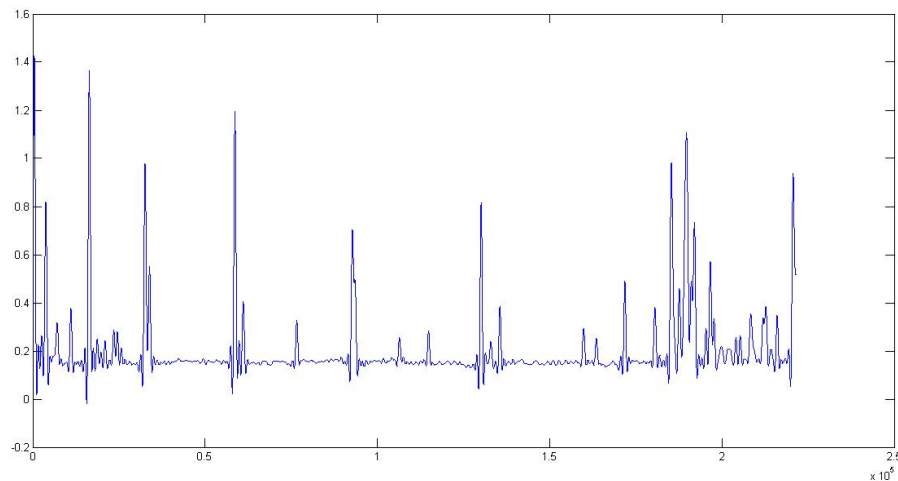


Figura 14: Filtrado

3.7. Extracción y Selección de Características

Una vez se ha llevado a cabo la etapa de preprocesado y se han eliminado aquellas partes de las señales que no aportan información al estudio, se extraen ciertas características con el propósito de representar los datos a través de otros valores más significativos que ayuden a la tarea de diferenciar entre ataques y no ataques. Aunque hay múltiples características que pueden ser extraídas, interesa seleccionar aquellas

que proporcionen mayor información y descartar aquellas que ayuden mínimamente a conseguir una respuesta correcta por parte del clasificador.

1. **Media:** se calcula el promedio de un conjunto de valores. Este tipo de parámetro es muy característico en distribuciones de probabilidad.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

2. **Desviación típica:** Una vez es conocido el valor promedio de un conjunto de datos, la desviación típica [15] nos da una medida de cuanto se dispersan los datos de ese valor promedio. Además, se puede deducir si un conjunto se aleja de su predicción ya que la desviación típica también puede ser interpretada como una medida de incertidumbre.

$$\sigma_s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (4)$$

3. **Máximo:** para facilitar la detección de picos, se obtiene el valor máximo de cada subconjunto de datos.
4. **Energía:** Obtener el valor de la energía procedente de la señal del ACC es importante para ser capaces de establecer un umbral para la detección de crisis. Se observa con mayor claridad la intensidad de los movimientos haciendo que la interpretación de los datos sea más fácil.

Como se muestra en la Figura 14, la magnitud de la aceleración varía según se encuentre el sensor en una situación de reposo o de actividad. El procedimiento que se ha seguido esta basado en la creación de una **ventana deslizante** durante la cual se extraen las características. Esta ventana se desplaza a lo largo de toda la señal con un solapamiento del 50 % que facilita la extracción de características durante periodos de cambios bruscos en los datos. La elección del tamaño de la ventana es en muchos casos el motivo de que un algoritmo de detección falle. Si el solapamiento de la ventana coincide con el final de una actividad y

el principio de otra, es posible que la extracción de características lleve a un error. Además, si escogemos un tamaño de ventana muy pequeño, puede que no nos de la información necesaria sobre los datos, por lo que determinar su longitud será un factor clave.

4. Procesado de las Señales

En este capítulo se explican todos los conceptos teóricos necesarios para la correcta interpretación del trabajo. Se hace una breve introducción al termino clasificador y se proponen varios modelos para el desarrollo de un algoritmo de detección. Además, se explican algunas medidas útiles para interpretar el rendimiento de los modelos propuestos.

4.1. Introducción

La evaluación de la actividad física que obtenemos a través de los sensores es posible gracias al uso de clasificadores.

Un clasificador [16] es un sistema utilizado para procesar fenómenos que han de ser descritos mediante estadísticas. Su objetivo es asignar una categoría concreta a un elemento de entrada. Generalmente un clasificador se implementa a través de una serie de entradas etiquetadas que permiten desarrollar un algoritmo que identifique las posibles salidas.

Se pueden construir una gran gama de clasificadores dependiendo del criterio que se quiera seguir y de los tipos de datos de los que se dispongan [17]. Por un lado, se encuentran los clasificadores supervisados y no supervisados. En el **aprendizaje supervisado** el conjunto de entrenamiento está etiquetado. Esto quiere decir que el clasificador sabe a priori a qué tipo de categoría pertenece cada una de las muestras que va a usar durante la fase de entrenamiento. Por el contrario, el **aprendizaje no supervisado** se basa en que el clasificador solo sabe el número de clases que debe identificar y de esta manera responde asignando etiquetas a las muestras de entrada.

Además, los clasificadores también pueden distinguirse entre secuencial y de aislamiento. El **secuencial** tiene en cuenta las clasificaciones que realizó en el pasado para inferir los resultados de las muestras futuras mientras que los de **aislamiento** tratan todas las muestras de manera independiente sin tener en cuenta la relación que puedan tener con las anteriores.

También, es posible hacer una segunda clasificación en función del método de trabajo que sigan los clasificadores:

- **Probabilísticos**

En los clasificadores probabilísticos la salida se basa en la probabilidad condicional $p(x|C_i)$ donde x es el vector de características de las entradas y C es el número de clases que se quieren identificar. En general, la probabilidad condicional es desconocida por lo que es necesario acudir a otros modelos como el Bayesiano o el Logístico para que nos permitan estimarla.

- **Geométricos**

Estos clasificadores construyen unos límites sobre las características de entrada para distinguir entre las diferentes clases. Durante el periodo de entrenamiento se realizan múltiples iteraciones para optimizar estos límites. Ejemplo de este tipo de clasificadores son las Redes Neuronales, las Maquinas de Vectores de Soporte o los Umbrales.

4.2. Algoritmos Propuestos

Uno de los aspectos más importantes a tener en cuenta a la hora de crear un sistema capaz de detectar cierta clase de evento es la elección del clasificador. A continuación se va a explicar la teoría relativa a los clasificadores usados en este proyecto.

4.2.1. Aproximación Mediante Umbral

La primera aproximación que se ha llevado a cabo para la detección de ataques de epilepsia es el uso de un umbral. Aunque no se presenta como un algoritmo de detección como tal, ayudará a determinar los valores de energía que presentan los ataques así como su duración.

Teniendo solo en cuenta la señal del ACC, se escogen varios pacientes al azar como conjunto de entrenamiento y se establecen varios umbrales de la siguiente manera:

1. Se elige un primer umbral basado en la energía media registrada de cada ataque. Aquellos picos que superen el umbral se catalogan como posibles ataques, tal como ilustra la Figura 15:

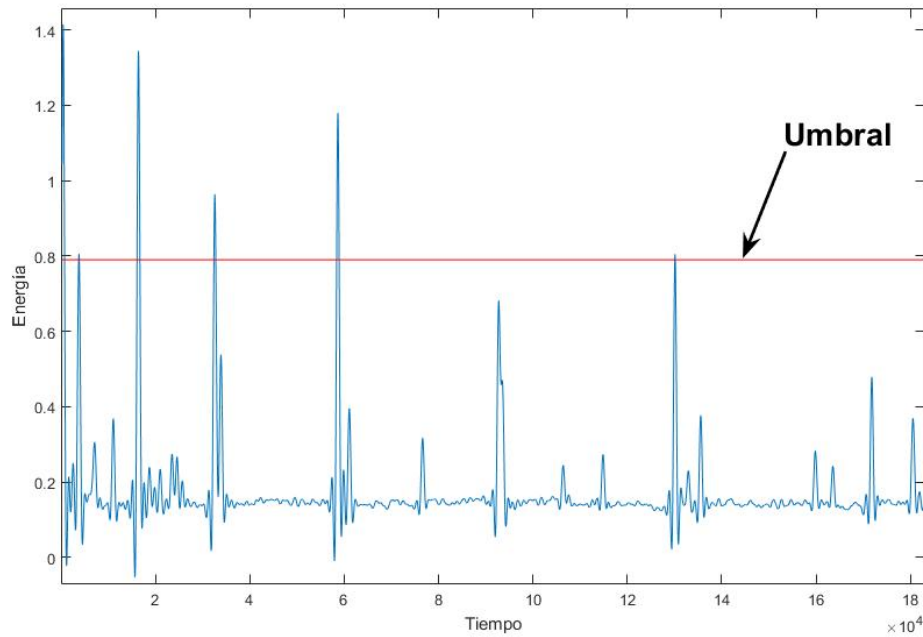


Figura 15: Aplicación de un umbral al acelerometro

2. A continuación, de todos aquellos picos que han superado el primer umbral, se descartan aquellos de poca duración. El objetivo es evitar clasificar como ataques aquellos movimientos que ha podido realizar el paciente durante la noche. Se medirá la duración de los picos que pasaron el primer umbral y se establece un segundo cuyo propósito es descartar los picos de corta duración.
3. A pesar de tener en cuenta la duración de los ataques, se ha propuesto un tercer umbral más alto que el primero, que nos servirá para catalogar directamente como ataques aquellos picos que lo superen sin tener en cuenta el punto 2.

El resultado final obtenido al aplicar este algoritmo se ilustra en la Figura 16. La principal desventaja de este enfoque es la potencial sensibilidad a las variaciones individuales y a la colocación precisa de los sensores.

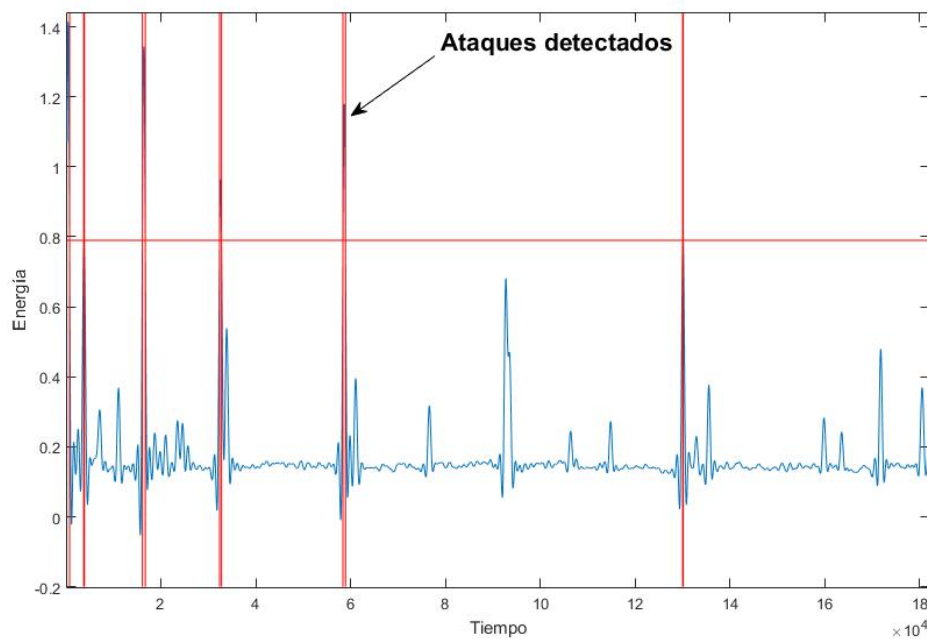


Figura 16: Detección de ataques a través de un umbral aplicado al acelerómetro.

4.2.2. Regresión Logística Binaria

La Regresión Logística (RL) [18] es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación clínica. Este modelo es utilizado para predecir el resultado de una variable dependiente en función de una o varias variables explicativas.

Los modelos de RL tienen dos finalidades definidas por sus características:

- Por un lado, permiten cuantificar cuál es la relación entre las variables dependientes y la variable a predecir.
- También permiten clasificar sucesos dentro de las categorías de la variable dependiente, según la probabilidad que esta tenga de pertenecer a cada una de esas categorías.

El objetivo de usar esta técnica será el de estimar la probabilidad de que ocurra un suceso y se modela a través de una función logística como la que se observa en la Figura 17.

Una variable binaria es aquella que puede adquirir dos posibles valores (ataque-no ataque), y tiene una especial importancia en teoría

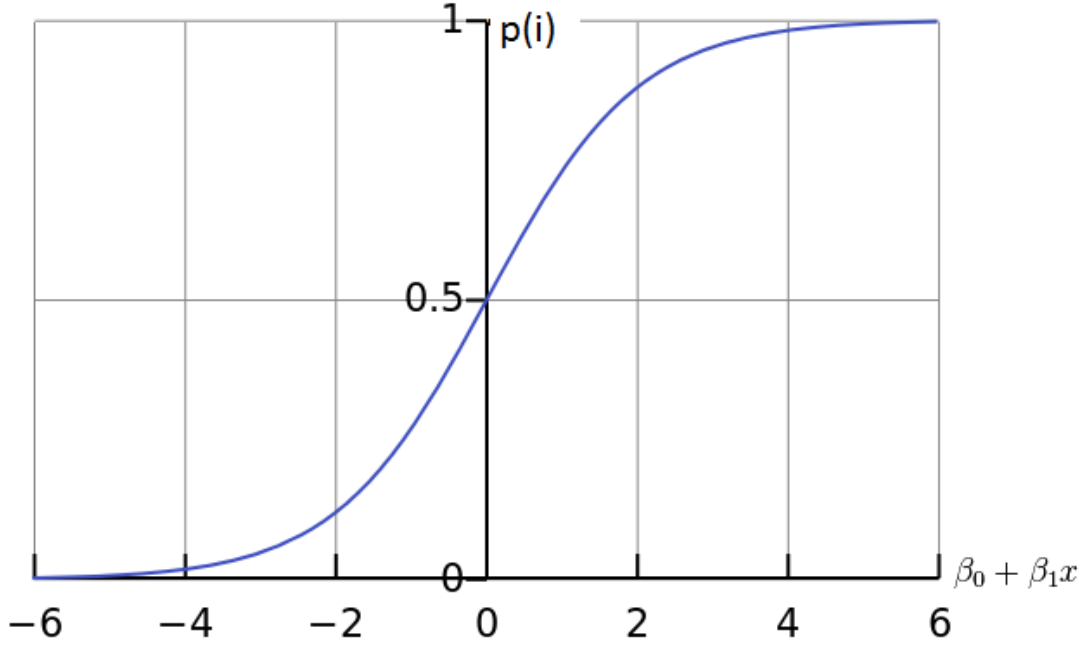


Figura 17: Función logística con $\beta_0 + \beta_1 x$ en el eje horizontal y p_i en el eje vertical.

de decisión. Cuando se pretende explicar el comportamiento de una variable binaria a través de un modelo de regresión, se utiliza un modelo de regresión lineal como el de la Formula 5, donde Y es la variable dependiente y X es la variable o conjunto de variables independientes.

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad (5)$$

Sin embargo, se ha demostrado que estos modelos presentan varios problemas cuando la variable dependiente es binaria por lo que usaremos modelos de regresión no binarios, como el modelo logístico. Este modelo tiene una formula dada por:

$$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \quad (6)$$

La estimación de los parametros *beta* se realiza a través de la función *mnrfit* de MATLAB, de la que obtenemos un vector o matriz, dependiendo de si solo tenemos una variable independiente o varias.

$$B = \text{mnrfit}(X, Y) \quad (7)$$

Para la aplicación de este método a la detección de ataques de epilepsia hemos llevado a cabo dos aproximaciones:

- Primero, se ha probado a aplicar este método teniendo solamente en cuenta el ACC. Además, de los 3 ejes del ACC, hemos tomado simplemente el valor absoluto, como podemos ver en la Formula 8.

$$Acc = \sqrt{Acc_x^2 + Acc_y^2 + Acc_z^2} \quad (8)$$

De esta manera, nos queda una única variable independiente X_1 igual al valor absoluto del ACC. El modelo de RL quedaría de la siguiente manera:

$$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1)} \quad (9)$$

La Figura 17 representa la relación entre el modelo de RL, expresado en la Formula 5 y la probabilidad de detección a través de una curva logística.

- Para la segunda aproximación que se ha llevado a cabo se ha incluido el valor absoluto del GYR, además de la variable X_1 que se aplicó en el primer punto. Ahora la variable independiente es colectiva y el modelo de regresión se expresa tal como aparece en la Formula 10.

$$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1 + \beta_2 X_2)} \quad (10)$$

El resultado de este modelo se trata como si fuese una hipótesis, donde un valor cercano a "0" representa una probabilidad baja de ataque, y un valor cercano a "1" representa alta probabilidad de ataque. Finalmente, elegiremos un umbral para el cual, todos aquellos resultados por encima de ese valor serán clasificados como ataques y los que no quedarán descartados.

El umbral escogido se basa en la experiencia obtenida de la aproximación que se llevó a cabo en el Apartado 4.2.1, dado que ya se conoce mejor la forma y características de los ataques.

4.2.3. Redes Neuronales

Las Redes Neuronales (RN) [16] son una técnica de procesamiento automático inspirado en la forma en la que funciona el sistema nervioso y como tal, intenta reproducir el comportamiento del cerebro.

Una RN se define como un conjunto de unidades de procesamiento llamadas neuronas, capaces de almacenar información y de conectarse entre ellas.

Arquitectura

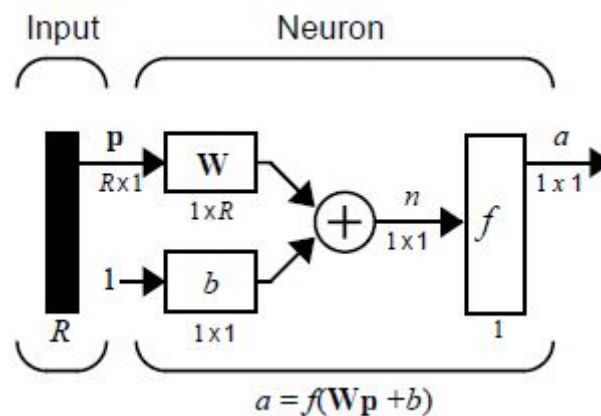


Figura 18: Neurona.

Las neuronas se organizan formando capas como la de la Figura 18. El vector de entrada p se representa como la barra negra vertical y tiene un tamaño de $R \times 1$ donde R es el número de muestras que tiene p .

$$p_1, p_2, \dots, p_R \quad (11)$$

Cada neurona recibe una serie de entradas provenientes del exterior o de sus interconexiones y emite una salida que será enviada a otras neuronas o al entorno.

La entrada p se transmite por una conexión que lo multiplica por un peso w formando el producto wp . A este producto se le añade el valor del sesgo, que simplemente se puede entender como un desplazamiento b de la función de transferencia. La suma $n = (wp + b)$ es la entrada de la función f de transferencia.

$$n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b \quad (12)$$

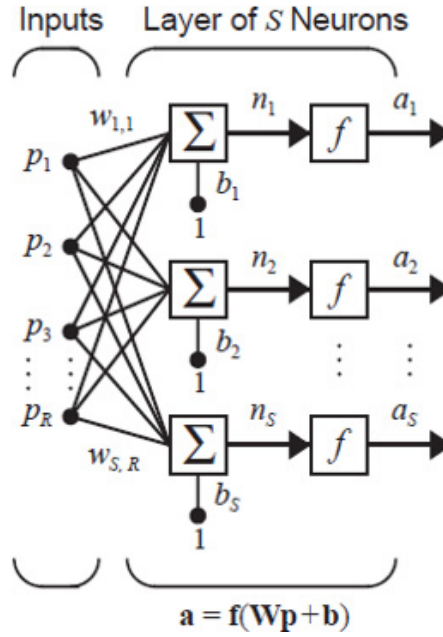


Figura 19: Modelo de Red Neuronal con una capa y múltiples neuronas.

En la Figura 19 podemos ver un ejemplo con S neuronas y R elementos de entrada. En este caso el vector p se conecta a cada capa de neuronas a través de la matriz de pesos W .

$$\begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{S,1} & w_{S,2} & \dots & w_{S,R} \end{pmatrix}$$

Las filas de la matriz de pesos indican la neurona a la que están destinados los pesos y las columnas la entrada a la que se le debe aplicar.

Sin embargo, las Redes Neuronales suelen contener más de una capa para optimizar el sistema tal como se representa en la Figura 20. De esta manera son más potentes y en cada una de las capas se pueden aplicar funciones diferentes que mejoren el resultado de la clasificación. En este tipo de redes las capas que contienen las salidas se denominan *capas de salida* y el resto *capas ocultas*.

En estadística, el riesgo empírico es un principio usado para imponer límites sobre el procesamiento de los algoritmos de aprendizaje.

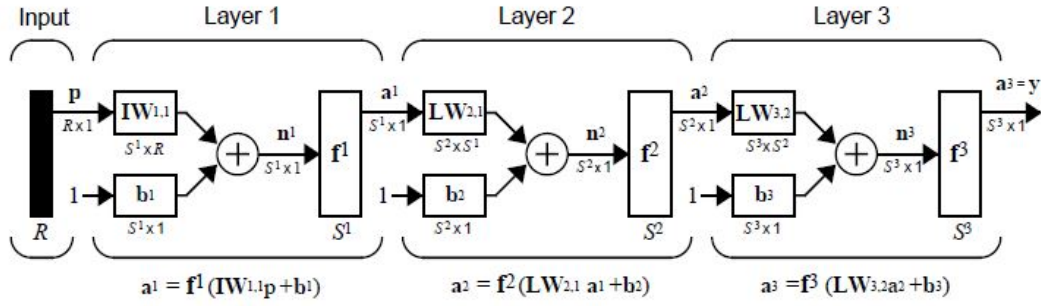


Figura 20: Red Neuronal con múltiples capas.

Cuando se dispone de unas muestras de entrenamiento y sus correspondientes etiquetas, lo que se busca es encontrar una función que las relacione de manera que se puedan deducir las etiquetas de las muestras. Sin embargo, esa función siempre es desconocida y lo que se formulan son hipótesis sobre las cuales se quiere minimizar el riesgo. Las RN se basan en este principio y se verá en el Apartado 4.2.4, que existen otros métodos, como las Máquinas de Vectores de Soporte, que se basan en la minimización del riesgo estructural.

Ventajas

Las RN son muy utilizadas en el campo de la medicina [19], [20]. Para su desarrollo se les provee de un conjunto de entrenamiento que consiste en patrones representativos de las distintas categorías que se van a clasificar y a partir de este proceso la maquina adquiere experiencia a través de la cual se deducen unas reglas que ayuden a decidir a qué clase se va a catalogar cada muestra.

Una de las mayores ventajas es que poseen un aprendizaje adaptativo, resultado del entrenamiento del sistema. Son capaces de organizar la información que reciben durante el entrenamiento y crear su propia estructura de datos. Además, tienen una alta tolerancia a fallos, es decir, que pese a recibir información con ruido no pierden su funcionalidad. Esto es así ya que la información que se halla distribuida en todas las neuronas tiene cierto grado de redundancia.

Finalmente, este tipo de sistemas de aprendizaje automático se computa de forma paralela lo que acorta la operación en tiempo real y deriva en un uso fácil y rápido en el que la red puede ser entrenada

y comprobada en poco tiempo y a bajo coste.

Aplicación en MATLAB

MATLAB proporciona al usuario una interfaz para facilitar el uso de las RN. Existen varios modelos que pueden utilizarse para desarrollar la red, como ilustra la Figura 21. Para la realización de este trabajo únicamente se ha utilizado el modelo destinado al reconocimiento de patrones.

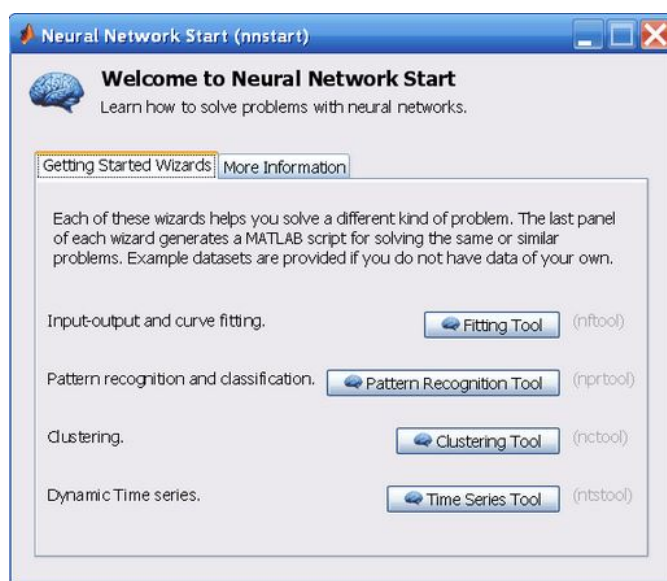


Figura 21: Interfaz de MATLAB.

El primer paso es la introducción de las características extraídas del conjunto de datos de entrenamiento así como de las etiquetas asociadas a esas características.

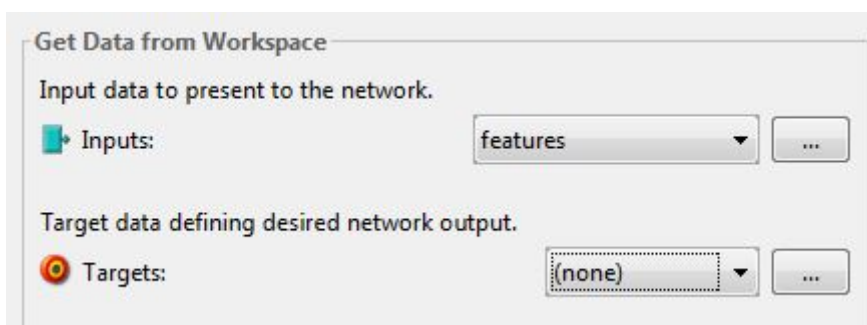


Figura 22: Interfaz de MATLAB.

De todo el conjunto de datos, se realiza una división aleatoria

entre conjunto de entrenamiento, validación y test. Los porcentajes de división de los datos se pueden modificar tal como aparece en la Figura 23.

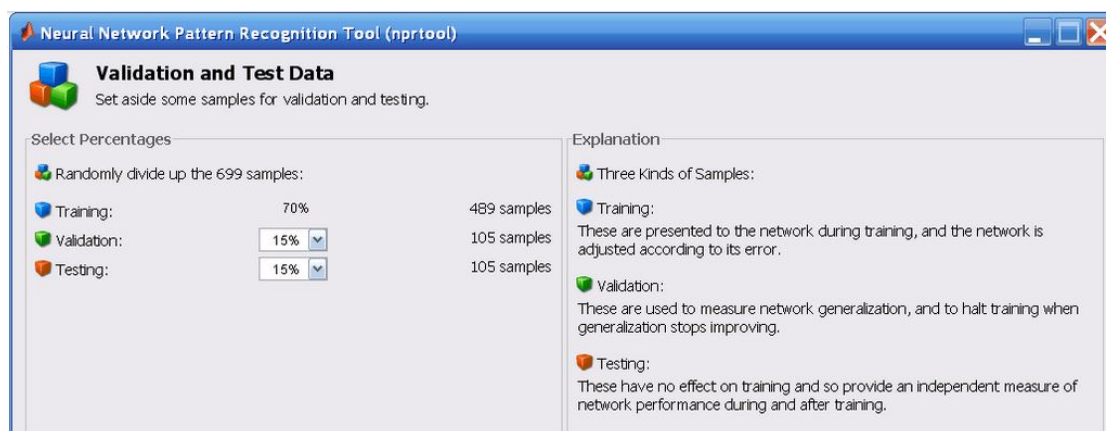


Figura 23: Interfaz de MATLAB.

El conjunto de entrenamiento será el que use la red neuronal para poder llevar a cabo el aprendizaje máquina. Al finalizar la etapa de entrenamiento, se validan los datos con el objetivo de comprobar la presencia de sobreentrenamiento y de seleccionar de manera óptima los parámetros que minimicen el error.

Finalmente, el algoritmo desarrollado por la RN se aplica sobre los datos de test. La interfaz permite visualizar el funcionamiento de la RN: por un lado, se puede comprobar cómo de exacto ha sido el modelo a través de las matrices de confusión. Un ejemplo se ilustra en la Figura 24. Se obtiene una matriz por cada etapa del modelo y su interpretación se basa en que cada columna representa el número de predicciones de cada clase y cada fila representa las instancias de la clase real.

Los elementos que pertenecen a la diagonal de la matriz representan los datos que han sido correctamente clasificados mientras que el resto de elementos representan los datos mal clasificados. Esta una manera fácil de ver si el sistema está funcionando correctamente o, si por el contrario, está confundiendo las clases.

Por otro lado, también puede visualizarse la curva ROC (acrónimo de Característica Operativa del Receptor), que representa la relación entre la sensibilidad frente a la especificidad. En la Figura 25 se

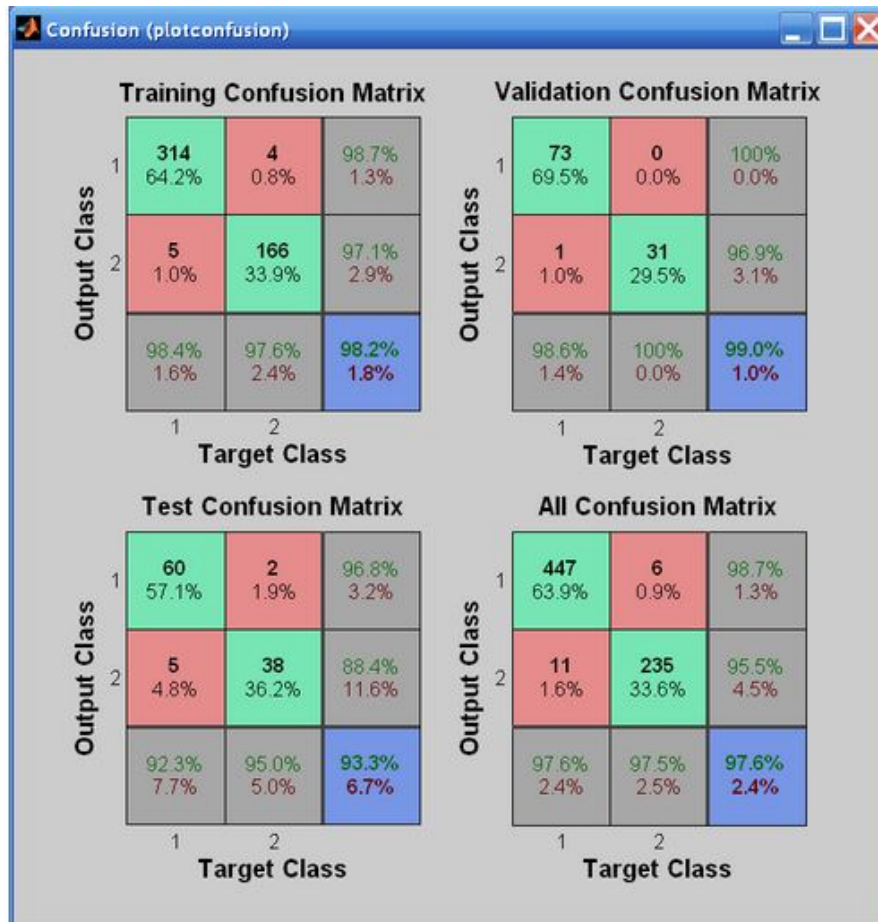


Figura 24: Interfaz de MATLAB.

representa un ejemplo, donde un valor en la curva cercano a la esquina superior izquierda significa una clasificación perfecta.

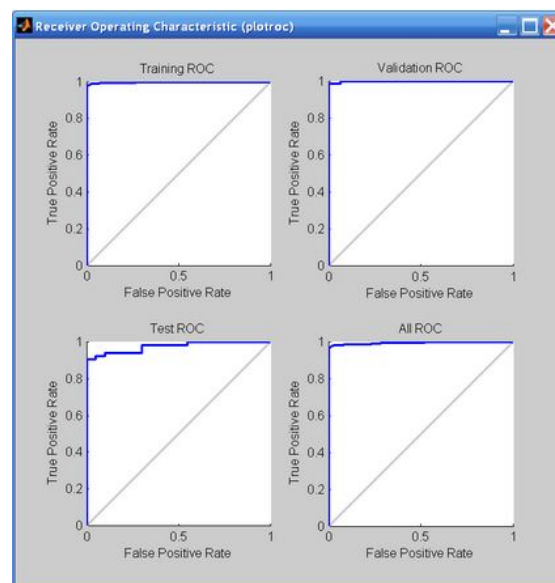


Figura 25: Interfaz de MATLAB.

4.2.4. Máquinas de Vectores de Soporte

Las Máquinas de Vectores de Soporte (SVM) son un conjunto de algoritmos basados en el aprendizaje supervisado y fueron desarrollados por V. Vapnik [21] a principios de los años 80 para resolver problemas de clasificación y reconocimiento de patrones.

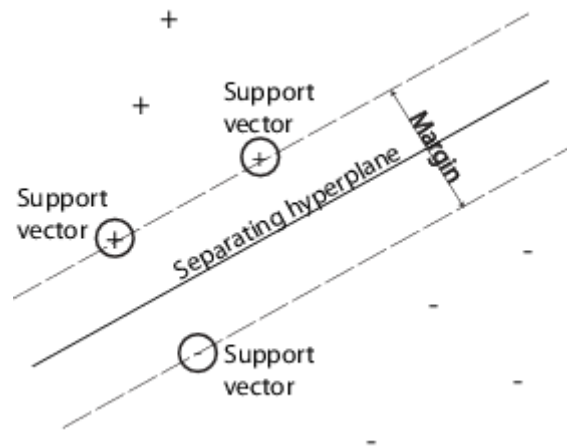


Figura 26: Hiperplano de las SVMs.

La teoría de las SVM se centra en mapear los puntos de entrada a un espacio de características de mayor dimensión con el fin de encontrar un hiperplano óptimo que los separe y maximice el margen entre las distintas categorías a clasificar.

En la literatura de las SVM se le denomina *atributo* a la variable predictora y *característica* al atributo extraído de los datos de entrada para definir el hiperplano. Se llama *vector de soporte* al vector formado por los puntos más cercanos al hiperplano.

A diferencia de las RN que se basaban en el principio de minimización del riesgo empírico (ERM), las SVM funcionan bajo el principio de minimización del riesgo estructural (SRM), consiguiendo alcanzar un límite superior al riesgo esperado con ERM, que simplemente actúa sobre los datos de entrenamiento. Algunos de los motivos por los que este método funciona mejor es que no padece de mínimos locales y solo depende de los datos que llevan mayor información, los vectores de soporte.

La formulación matemática necesaria para llevar a cabo este modelo dependerá de la naturaleza de los datos, ya sea lineal o no lineal. Para el desarrollo de este proyecto se han usado SVM lineales ya que se trata de un caso de clasificación bidimensional.

Modelos lineales de vectores soporte

Para el caso en que el conjunto de datos $\{x_1, y_1\}, \dots, \{x_n, y_n\}$ sea linealmente separable, las SVM forman hiperplanos que separen los datos de entrada en dos subgrupos de etiquetas denotadas por $Y = \{-1, 1\}$. A pesar de encontrar varios hiperplanos posibles, solo existe un hiperplano de separación óptimo y será aquel en el que la distancia entre el hiperplano y el valor de entrada sea la máxima posible.

$$\pi : w \cdot x + b = 0 \quad (13)$$

El problema de la optimización se plantea de la siguiente manera: fijado un hiperplano, es posible recalcular los valores de los parámetros w y b , para todo valor de x_i de tal forma que:

$$\pi_1 : w \cdot x_i + b \geq 1 \text{ para } y_i = +1 \quad (14)$$

$$\pi_2 : w \cdot x_i + b \leq -1 \text{ para } y_i = -1 \quad (15)$$

El valor del margen se expresa de la siguiente forma:

$$\frac{2}{\|W\|} \quad (16)$$

El objetivo final es maximizar la expresión de la Formula 16 siempre sujeta a las Expresiones 14 y 15. Así se consigue que la separación entre los vectores y el hiperplano sea la unidad.

La solución para el caso bidimensional se puede interpretar gráficamente a través de la Figura 27. Una característica muy importante de las SVM es que si se añade o elimina cualquier vector que cumpla la desigualdad estricta de la Fórmula 13, la solución al problema de la optimización no se ve afectada mientras que si se encuentra un vector localizado entre ambos planos π_1 y π_2 , la solución cambia totalmente.

Sin embargo el procedimiento matemático se puede transformar para que sea más fácil de manejar haciendo uso de los multiplicadores

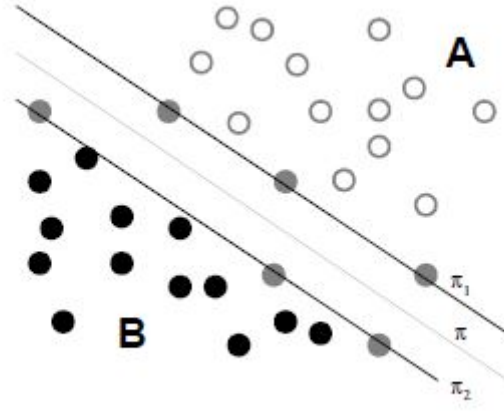


Figura 27: Conjunto e hiperplano separador. Los puntos huecos representan los vectores con etiqueta $y = 1$ y los rellenos $y = -1$.

de Lagrange. Esta es una alternativa común para encontrar máximos o mínimos de funciones sujetas a restricciones. En este caso el valor del margen se expresa como $\|W\|^2$ y el objetivo es minimizar su valor con lo que el problema se reduce a la siguiente expresión:

$$L_p = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^l \alpha_i \quad (17)$$

Haciendo que los gradientes de L_p respecto a w y b sean cero, se obtienen las siguientes expresiones:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (18)$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (19)$$

Y reemplazando en L_p se obtiene el problema dual:

$$L_D = \sum_{i=1}^l \alpha_i + \sum_{i=1, j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \quad (20)$$

Finalmente la forma para optimizar se basa en maximizar la Expresión 20 sujeta a las Expresiones 18 y 19.

Los vectores del conjunto de entrenamiento que proporcionan un multiplicador $\alpha_i \geq 0$ se denominan **vectores soporte** y se encuentran sobre uno de los hiperplanos π_1 y π_2 . Para este tipo de algoritmos, estos vectores son muy importantes puesto que son los que dan la aproximación del problema. Si el resto de elementos del conjunto de entrenamiento fuesen eliminados y el problema de optimización cambiase, se seguirían obteniendo los mismos hiperplano separadores.

Hasta ahora, se ha visto el desarrollo matemático al problema de clasificación con SVM y datos linealmente separables. En la práctica, lo habitual es encontrarse con que los datos no se pueden separar linealmente. Ante estos casos, el problema de optimización no encuentra una solución ya que la función objetivo 17 aumenta arbitrariamente. Partiendo del planteamiento lineal, se añade una variable ε de holgura en las restricciones y se plantea un nuevo conjunto:

$$w \cdot x_i + b \geq +1 - \varepsilon_i \text{ para } y_i = +1 \quad (21)$$

$$w \cdot x_i + b \leq -1 + \varepsilon_i \text{ para } y_i = -1 \quad (22)$$

$$\varepsilon_i \geq 0 \quad \forall i = 1, \dots, n \quad (23)$$

Para poder ser capaces de separar los datos se hace uso de una *función kernel* [22]. La idea principal es la de proyectar los puntos de datos a un espacio dimensional mayor con el objetivo de mejorar su separación. Además del kernel, se hace uso de un parametro C que determina el intercambio entre los dos objetivos del algoritmo: maximizar el margen y minimizar los errores de entrenamiento. Para que se produzca una clasificación errónea, es necesario que el valor de la holgura ε_i sea superior a la unidad.

Ventajas de las Máquinas de Vectores de Soporte

Se han encontrado algunas características que aportan a las SVM ciertas ventajas con respecto a otros algoritmos de clasificación. Una de ellas es que pertenecen al aprendizaje automático o estadístico cuyo objetivo es que la máquina aprenda a través de ejemplos y sea capaz de obtener resultados correctos para datos de entrada nuevos. Existen pocos parámetros que se deban ajustar para llevar a cabo el modelo ya que solo depende de los datos con mayor información como los vectores

de soporte.

Sin embargo, la ventaja más importante de las SVM es que aplican un modelo inductivo a través de la minimización del riesgo estructural, además del uso de una función núcleo que aporta al algoritmo una gran capacidad de generalización, incluso si el conjunto de entrenamiento es pequeño.

Problemas con las Máquinas de Vectores de Soporte

Cuando se trabaja con las SVM, se pueden encontrar varios errores de entrenamiento. Si el algoritmo funcionase de manera ideal, se debería encontrar un hiperplano óptimo que separase completamente todos los datos en dos subcategorías. Esto no es así, ya que un modelo no puede ser generalizado para todo tipo de datos. Cuando la complejidad del clasificador aumenta, la salida se vuelve muy sensible a pequeñas variaciones en los valores de entrada, por lo que aunque se consiga maximizar el valor del margen entre las dos categorías, es probable que el algoritmo sufra **sobreajuste**. Para poder introducir cierta flexibilidad al problema, se añade un parámetro C encargado de controlar la compensación entre los errores de entrenamiento y los márgenes. Como resultado, se crea un margen blando que, aunque permite algunos errores en la clasificación, también los penaliza.

Otro de los problemas que se pueden encontrar trabajando con SVM es el **sobreentrenamiento**. Este fenómeno ocurre cuando se han aprendido tan bien los datos de entrenamiento que el clasificador es incapaz de catalogar correctamente los nuevos datos. La porción n de los datos no conocidos que se clasificará de manera errónea está limitada por la siguiente expresión:

$$n = \frac{\text{num. de vectores de soporte}}{\text{num. de ejemplos de entrenamiento}} \quad (24)$$

Una mala clasificación también puede deberse a que el algoritmo no ha aprendido bien las características del conjunto de entrenamiento. Por ello es importante ajustar bien el modelo. Aunque en determinados casos las SVM demuestren ser superiores a las RN, SVM tiene el

inconveniente de la selección del kernel y dificultades al realizar una clasificación sobre un conjunto muy amplio de datos de entrenamiento. De estos factores dependerá la exactitud y precisión de la predicción.

Aplicación en MATLAB

Para poder implementar las SVM, MATLAB cuenta con una serie de funciones predefinidas. Como con cualquier modelo de entrenamiento, primero se ha de entrenar la máquina y una vez se ha hecho, se validan los resultados.

Con el objetivo de entrenar la máquina, se ha usado la Función *fitcsvm*, cuya sintaxis es:

```
SVMModel = fitcsvm(X,Y,'KernelFunction','rbf','Standardize',true,'ClassNames',{'negClass','posClass'});
```

En esta Función se introducen las características extraídas de los datos de entrada y sus correspondientes etiquetas. Además, para obtener una precisión óptima, se pueden usar varios núcleos distintos, que se deben especificar como parámetro. Para casos de predicción de dos tipos de eventos, el valor por defecto es el kernel '*linear*', que separa los datos a través de un hiperplano. La elección del kernel adecuado es un paso de gran importancia para el correcto funcionamiento del algoritmo. A través del parametro *ClassNames* podemos especificar el nombre de la clase que estamos intentando identificar. Se hace una distinción entre clase negativa y positiva. Para el desarrollo de este trabajo identificaremos la clase negativa como que 'no hay ataque' y, por el contrario, se identificará la clase positiva como que 'si hay ataque'. La función devuelve como resultado un modelo de entrenamiento que contiene los parametros de optimización que se usarán para clasificar posteriormente los datos nuevos.

Para llevar a cabo la fase de clasificación se usa la Función *predict* de la siguiente manera:

```
[label,score] = predict(SVMModel,newX);
```

En ella se incluyen tanto el modelo que se obtuvo en la fase de entrenamiento, como las características extraídas de los nuevos datos. Finalmente, se obtiene un vector *label* que contiene los resultados de

la clasificación y se interpreta como la probabilidad de que el suceso a detectar ocurra. *Scores* es una matriz de $n \times 2$ que representa las salidas blandas. Por un lado, en la primera columna contiene la probabilidad de que el suceso negativo ocurra mientras que en la segunda columna contiene la probabilidad de que el suceso positivo ocurra.

Comparativa entre Máquinas de Vectores de Soporte y Redes Neuronales

Tanto las RN como las máquinas de vectores de soporte son dos estrategias muy utilizadas para el aprendizaje supervisado y la clasificación. Determinar cuál de ellas funciona mejor depende de los datos que se estén usando. En general, ambos son muy buenos métodos y tienen un funcionamiento notable en problemas típicos. Los algoritmos de ambos métodos pueden formularse a través del aprendizaje máquina: El desarrollo del clasificador consiste, como ya se ha mencionado anteriormente, en dos etapas: entrenamiento y test. Durante la primera etapa se presentan las muestras etiquetadas para que la máquina sea capaz de aprender y finalmente se comprueba el rendimiento del clasificador durante la etapa de test.

Sin embargo, hay ciertos factores que las diferencian y hacen que su rendimiento no sea el mismo. Las SVM son computacionalmente más rápidas siempre que la elección del kernel sea la correcta mientras que las ANNs son más lentas durante la etapa de entrenamiento aunque mejoran su tiempo de cómputo en la clasificación.

Uno de los beneficios que poseen las RN es que su tamaño es fijo: cuando se crea una red neuronal se especifican el número de capas ocultas que se desean y ese parámetro se mantiene durante todo el proceso de entrenamiento mientras que para las SVM se diseña la función kernel y el parámetro de coste pero el número de vectores soporte que pueden encontrarse es variable. En el peor de los casos, ese número es exactamente igual al número de muestras de entrenamiento. A pesar de ello, el procesamiento en SVM con cientos de funciones es algo normal y el entrenamiento es muy eficiente mientras que para las RN el entrenamiento se hace costoso.

Las RN están basadas en la minimización del error cuadrático me-

dio, por lo que el aprendizaje puede finalizar debido a dos posibles motivos: porque se ha encontrado la cota de error buscada o porque se han realizado el número máximo de iteraciones. Si no se alcanza la cota de error, el modelo o los datos de entrenamiento deben cambiarse. Aunque con este tipo de algoritmos se pueden obtener soluciones múltiples, no se asegura que sea la óptima. Una de las debilidades de las RN es que a menudo convergen en mínimos locales en lugar de mínimos globales, lo que significa que carecen de una visión de todo el conjunto de datos mientras que con las SVM el espacio de búsqueda tiene únicamente un mínimo global, lo que implica que la solución es única, y se trata de la solución óptima que maximiza el margen. Además, tal como ocurre con las SVM, las RN a menudo también sufren de sobreajuste si el conjunto de datos es demasiado grande. Esto deriva en que para cualquier patrón, las RN podrían considerar el ruido como parte del patrón.

4.3. Medidas de rendimiento

En el Apartado 5 se evalúa el funcionamiento de los algoritmos propuestos para la detección de ataques de epilepsia que se han comentado a lo largo de este capítulo. Para ello, se hace uso de la matriz de confusión, la cual contiene una serie de conceptos que se explicarán a continuación:

La Figura 28 muestra los distintos conceptos que se van a tener en cuenta: hablaremos de *verdadero positivo* cuando el algoritmo detecte un ataque y este, efectivamente, haya ocurrido. Sin embargo, de no ser así, hablaremos de un *falso positivo*. Por el contrario, cuando el algoritmo no sea capaz de detectar el ataque y este haya ocurrido, se producirá lo que se conoce como *falso negativo*, mientras que si no se detectan ataques que, efectivamente, no han ocurrido, se trata de *verdadero negativo*.

RESULTADOS REALES DE LOS ATAQUES			
R E S U L T A D O S D E T E S T		<u>Presencia de Ataque</u>	<u>Ausencia de Ataque</u>
	<u>Positivo</u>	Verdadero Positivo <i>VP</i>	Falso Positivo <i>FP</i>
	<u>Negativo</u>	Falso Negativo <i>FN</i>	Verdadero Negativo <i>VN</i>

Figura 28: Matriz de confusión

Sensibilidad

Es una medida de la probabilidad de clasificar de manera correcta a un sujeto enfermo. Representa la capacidad del algoritmo para detectar los ataques de epilepsia [23]. Se obtiene a través de la siguiente relación:

$$S = \frac{VP}{VP + FN} \quad (25)$$

donde VP representa el número de verdaderos positivos y FN representa los falsos negativos. La sensibilidad también se denomina como *fracción de verdaderos positivos (FVP)*.

Especificidad

Al contrario que la sensibilidad, la especificidad es la probabilidad de clasificar correctamente a un sujeto sano. Representa la probabilidad de que el algoritmo detecte ataques sobre pacientes que no han sufrido ninguna crisis. Se obtiene a partir de la siguiente expresión:

$$E = \frac{VN}{VN + FP} \quad (26)$$

donde VN representa el numero de verdaderos negativos y FP , los

falsos positivos. La especificidad también se denomina *fracción de verdaderos negativos* (*FVN*).

Idealmente, se busca que los resultados siempre tengan altos valores de sensibilidad y especificidad aunque no siempre es posible. Para evaluar la efectividad del sistema, ambos valores se relacionan a través de la curvas ROC.

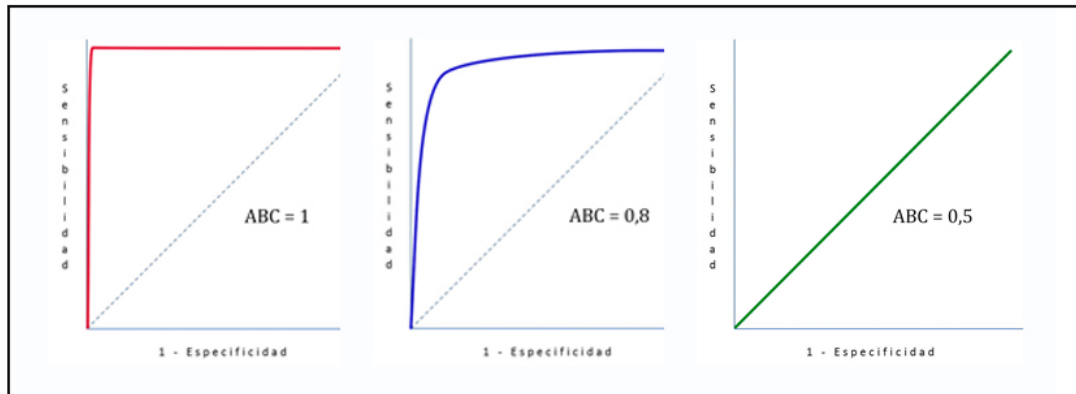


Figura 29: Curvas ROC

En ellas se representa la sensibilidad frente al valor obtenido de $1 - \text{especificidad}$, obteniéndose una curva cuyo área determina la efectividad del algoritmo propuesto. Si el área bajo la curva alcanza un valor cercano a 1, como en la gráfica de la izquierda, significa que el algoritmo diagnostica correctamente según las especificaciones deseadas. Por el contrario, si el área bajo la curva tiene un valor alrededor del 0,5, como se ilustra en la parte derecha de la Figura 29, se deduce que el detector no es capaz de discriminar entre la existencia de un ataque (valor positivo) o de que no lo haya (falso positivo). Entre ambos extremos, se pueden modificar los parámetros del modelo para conseguir que el algoritmo alcance los valores de sensibilidad y especificidad máximos. En la curva central de la Figura 29 podemos ver un ejemplo de una curva ROC que alcanza un valor de 0.8. En diagnósticos médicos, este es el valor mínimo que se debe adquirir para considerar que el algoritmo tiene viabilidad.

Precisión

La precisión es la relación entre los valores verdaderos positivos obtenidos y el total de valores positivos, incluyendo aquellos que se

han clasificado de manera incorrecta:

$$P = \frac{VP}{VP + FP} \quad (27)$$

donde VP es el número de verdaderos positivos y FP el número de falsos positivos. Da una medida acerca de lo relevantes que son los datos que se extraen.

Valor-F

El valor-F es una medida que combina ambos conceptos de sensibilidad y precisión con el fin de determinar el acierto obtenido durante la fase de test de un algoritmo. Se obtiene según la expresión 28:

$$F = 2 * \frac{P * S}{P + S} \quad (28)$$

donde P y S son la precisión y la sensibilidad respectivamente.

5. Resultados

Para comprobar la efectividad de los métodos propuestos y realizar los ajustes necesarios para obtener mejores resultados, se realizan una serie de pruebas, que ayuden a asegurar la calidad de los algoritmos clasificadores. En este apartado se mostrarán los resultados obtenidos al aplicar los distintos métodos de detección automática que se presentaron en el Apartado 4.

Sin embargo, durante la fase de pruebas se han encontrado pacientes mal etiquetados. Para evitar que los datos de esos pacientes dificulten el entrenamiento de los clasificadores y se obtengan resultados erróneos, se ha decidido descartarlos del estudio.

5.1. Regresión Logística Binaria

La aplicación del modelo de RL se ha llevado a cabo de dos formas: en la primera, se ha tomado unicamente el valor absoluto del ACC como entrada. En la segunda, se ha trabajado además con los datos procedentes del GYR.

En ambas aplicaciones, la extracción de características (Apartado 3.7) no es necesaria ya que se trabaja con el valor absoluto de ambos módulos del sensor.

Acelerómetro

La primera aplicación del modelo de RL que se ha llevado a cabo se ha realizado utilizando el valor absoluto del ACC. Se ha escogido un conjunto de pacientes significativo que representase tanto periodos de crisis como estados de reposo para favorecer la fase de entrenamiento.

Se etiquetan los datos procedentes del ACC y se obtienen los parámetros *beta* para aplicarlos sobre la Formula 9.

$$B = \begin{pmatrix} 15,3660 \\ -19,2036 \end{pmatrix}$$

El tiempo de procesado de la Función *mrnfit* es de 24 minutos y 12 segundos, lo que hace que sea el método computacionalmente más lento de todos los propuestos.

Aplicamos los parámetros *beta* sobre el conjunto de test obteniendo la siguiente clasificación de las muestras:

	Presencia de Ataque	Ausencia de Ataque
Positivo	5032	0
Negativo	250	1716285

Tabla 2: Clasificación de las muestras de test utilizando Regresión Logística: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.

Dada la clasificación de las muestras de la Tabla 2 obtenemos las siguientes medidas de rendimiento:

1. Sensibilidad = 95.25 %.
2. Especificidad = 100 %.
3. Precisión = 100 %.
4. Medida F = 97.57 %.

El algoritmo detecta todos los ataques existentes, pero al tratar todas las muestras de manera independiente, la sensibilidad se ve disminuida debido a que el inicio de las crisis a veces es marcada como un falso negativo, es decir, que el paciente ha empezado a sufrir una crisis pero el algoritmo no es capaz de detectarla. Sin embargo, se carece de falsos positivos, ya que en ninguna ocasión se detectan ataques durante periodos de reposo.

Acelerómetro y Giróscopo

La segunda aplicación del modelo se realiza utilizando el ACC de la misma manera que en el caso anterior, pero ahora se añade el GYR, con el propósito de comprobar si añadir información adicional mejora

los resultados del algoritmo. Además, se intercambian conjunto de entrenamiento y test con respecto al caso anterior.

La solución del modelo se obtiene según la Fórmula 10, para la cual se necesitan 3 parámetros *beta* que se obtendrán con la función *mnrfit* como ya se hizo previamente.

$$B = \begin{pmatrix} 34,5185 \\ -43,4439 \\ -0,0010 \end{pmatrix}$$

El tiempo de procesamiento de los parámetros *beta* vuelve a ser alto, en este caso de 36 minutos y 57 segundos.

Al aplicar los parámetros *beta*, la clasificación de las muestras para el conjunto de test es la siguiente:

	Presencia de Ataque	Ausencia de Ataque
Positivo	7485	54
Negativo	265	1529014

Tabla 3: Clasificación de las muestras de test utilizando Regresión Logística: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.

Las medidas de rendimiento para la clasificación obtenida son:

1. Sensibilidad = 96.58 %.
2. Especificidad = 99 %.
3. Precisión = 99.28 %.
4. Medida F = 97.91 %.

A pesar de haber cambiado el conjunto de test y de entrenamiento, los resultados son muy parecidos al caso en el que se usaba únicamente la señal del ACC. Añadir la señal obtenida del GYR no le aporta al sistema una información adicional relevante para mejorar sus características y el resultado. Si se tienen en cuenta directamente los valores de los parámetros *beta* que se obtuvieron para este modelo, se puede

observar que el parámetro aplicado sobre la señal del GYR es prácticamente nulo, lo que le dice al algoritmo que este no aporta información relevante para la clasificación.

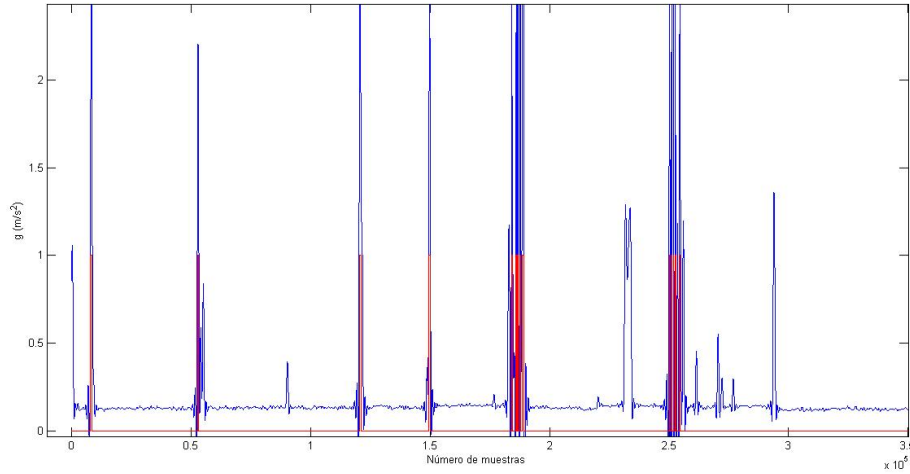


Figura 30: Ejemplo de detección con Regresión Logística.

En la Figura 30 se puede ver un ejemplo de los resultados obtenidos con RL. Se detectan todos los ataques presentes pero si, además, se observa con más detalle, se puede observar que en determinados ataques, la detección (identificada con líneas rojas) empieza algo después de que la crisis comience. Como ya se ha mencionado con anterioridad, esto se debe a que el algoritmo trata las muestras independientemente y no tiene en cuenta valores pasados para predecir futuros ataques.

5.2. Redes Neuronales

La implementación de las RN se ha realizado sobre los datos obtenidos del ACC. En este modelo, los datos serán representados a través de otras características más significativas que favorezcan el entrenamiento por lo que tras la etapa de pre-procesado, se extraen las siguientes características del ACC:

- Energía
- Máximo + Desviación Típica

La optimización del algoritmo se realiza teniendo en cuenta el valor de la ventana deslizante sobre la cual se extraen dichas características.

La variación del valor de la ventana determinará el éxito del funcionamiento de las RN.

En cuanto al tiempo de computación, las RN son lentas, pero no llegan a valores tan altos como el tiempo alcanzado por el modelo de RL. El procesamiento de las tres fases de entrenamiento, validación y test alcanza en torno al minuto y medio.

Ventana = 20 muestras

La primera aproximación realizada se hace utilizando una ventana deslizante de 20 muestras. La frecuencia de muestreo del sensor es de 8 Hz por lo que se consideran 2 segundos y medio de datos por cada ventana. Los resultados de las RN se presentan a través de la Matriz de Confusión de la Figura 31

1. Sensibilidad = 69 %.
2. Especificidad = 99.94 %.
3. Precisión = 82.67 %.
4. Medida F = 75.21 %.

El número de muestras que tienen que etiquetarse es muy elevado por lo que la precisión del algoritmo se ve afectada. Además, la división que se realiza sobre los datos para entrenamiento y test es aleatoria. De un entrenamiento a otro los datos pueden variar mucho. En este caso, la división de los datos de entrenamiento no ha sido la mejor, y el algoritmo no ha sido capaz de aprender bien las características extraídas del ACC. Por esta razón, la sensibilidad disminuye.

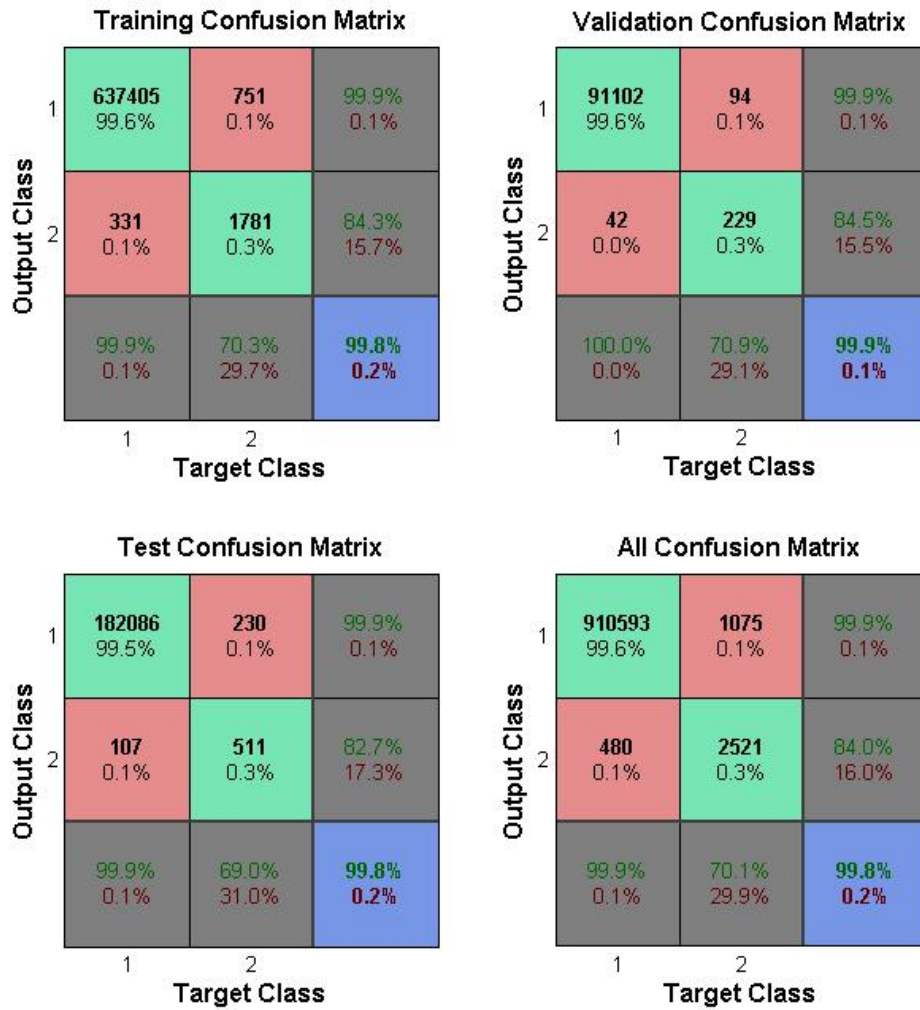


Figura 31: Matriz de confusión para una Red Neuronal con ventana deslizante de 20 muestras.

Ventana = 100 muestras

Para la segunda prueba realizada con RN se utiliza una ventana deslizante de 100 muestras. Se comprueba que para este caso los resultados son notablemente mejores. En la Figura 32 se ilustra la Matriz de Confusión obtenida.

1. Sensibilidad = 96.3 %.
2. Especificidad = 100 %.
3. Precisión = 95.1 %.



Figura 32: Matriz de confusión para una Red Neuronal con ventana deslizando de 100 muestras.

4. Medida $F = 95.69\%$.

La gran diferencia entre los resultados obtenidos con la ventana de 20 muestras y la de 100 resalta la importancia de elegir adecuadamente su tamaño. La correcta extracción de las características es clave para que el entrenamiento de cualquier modelo sea el adecuado.

De nuevo, debido a la aleatoriedad en la división de los datos para la fase de entrenamiento, test y validación, los resultados de un entrenamiento a otro pueden cambiar drásticamente. Aunque en determinados casos se alcancen valores muy altos en las medidas de rendimiento con las RN, esa aleatoriedad no beneficia al modelo

para nuestro caso específico. Ya que en los datos reales de los que se dispone no se registran demasiadas crisis, la división del conjunto de entrenamiento es muy importante y se debe hacer de la mejor manera, seleccionando los ataques más significativos para que aporten al modelo la mayor cantidad de información posible.

5.3. Máquinas de Vectores de Soporte

Por último, se mostrarán los resultados del método basado en SVM. Para el conjunto de datos seleccionado, se obtiene un modelo de entrenamiento a través de la función *fitcsvm*, como ya se explicó en el apartado 4.2.4. Las características extraídas del ACC son de nuevo:

- Energía
- Máximo + desviación típica

El tiempo de procesado de dicha función es muy bajo, de apenas segundos. Comparado este método es mucho más rápido.

Ventana = 20 muestras

La primera prueba del método se realiza sobre una ventana deslizante de 20 muestras. Esto nos da mayor detalle de la información extraída. Los resultados mostrados en la Tabla 4 son muy satisfactorios a pesar de haberse detectado un falso positivo. Todas las crisis existentes son identificadas correctamente aunque persiste el problema de la independencia de las muestras y por lo tanto, las crisis son detectadas cuando alcanzas un valor de energía alto. La influencia de muestras pasadas para deducir que se aproxima una crisis no tiene posible implementación con este modelo.

	Presencia de Ataque	Ausencia de Ataque
Positivo	826	1
Negativo	10	164861

Tabla 4: Clasificación de las muestras de test utilizando SVM: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.

Las medidas de precisión correspondientes a dicha clasificación son:

1. Sensibilidad = 98.99 %.
2. Especificidad = 99.9 %.
3. Precisión = 99.87 %.
4. Medida F = 99.3 %.

Ventana = 100 muestras

Se modifica el tamaño de la ventana para ver como afecta a la efectividad del modelo basado en SVM. La siguiente prueba realizada se hace con una ventana de 100 muestras, lo que conlleva un solapamiento de 50 muestras entre desplazamientos anteriores y posteriores. A continuación se muestra la Tabla de clasificación del conjunto de test:

	Presencia de Ataque	Ausencia de Ataque
Positivo	203	0
Negativo	70	40288

Tabla 5: Clasificación de las muestras de test utilizando SVM: {Verdadero Positivo, Falso Positivo, Verdadero Negativo y Falso Negativo}.

Las medidas de precisión correspondientes a dicha clasificación son:

1. Sensibilidad = 74.35 %.
2. Especificidad = 100 %.
3. Precisión = 100 %.
4. Medida F = 85.28 %.

La sensibilidad se ve reducida puesto que, al aumentar la ventana, se extraen características de un conjunto más amplio y los valores obtenidos de la energía y el máximo se suavizan. Se pierde sensibilidad frente a cambios bruscos en la señal y por esa razón aumenta el número de falsos negativos.

5.4. Resumen y Comparativa de los Modelos Propuestos

Una vez comentados los resultados de todos los modelos propuestos, la Tabla 6 facilita la comparación entre los distintos clasificadores:

Modelo	Implementación	Tiempo de Procesado	P	E	S	F
RL	ACC	Alto	1	1	0.952	0.975
	ACC+GYR	Muy Alto	0.992	0.999	0.9658	0.979
RN	ventana = 20	Medio	0.826	0.999	0.690	0.752
	ventana = 100	Medio	0.951	1	0.963	0.956
SVM	ventana = 20	Bajo	0.998	0.999	0.989	0.993
	ventana = 100	Bajo	1	1	0.753	0.852

Tabla 6: Resumen de Resultados

En cuanto al tiempo de procesado, las SVM son el modelo más rápido. Aunque las RN tienen un tiempo de procesado superior, distan bastante de alcanzado por la RL, por lo que de todas maneras las RN pueden considerarse un modelo de clasificación rápido.

Las medidas de rendimiento obtenidas con RL presentan un porcentaje más alto de acierto que el resto de modelos propuestos, aunque un gran número de muestras son clasificadas como falsos negativos. El objetivo de este proyecto es desarrollar un sistema de clasificación que sirva como complemento al diagnóstico de los neurólogos, por lo que el número de falsos positivos en este caso no es demasiado relevante. Sin embargo, los falsos negativos no son deseables puesto que de cara al diagnóstico, es menos perjudicial detectar una crisis de manera errónea, que pasarla por alto.

Ya se mencionó con anterioridad que las RN, aunque alcancen en ocasiones una clasificación muy precisa, varían mucho de un entrenamiento a otro debido a que la división del conjunto de entrenamiento y test se hace de manera aleatoria. Para realizar este proyecto, se dispone de datos reales recolectados de pacientes que fueron admitidos para monitorizarlos con fines de investigación. Los datos recolectados registran pocas crisis por lo que la elección del conjunto de entrenamiento es extremadamente importante y determina la efectividad del algoritmo. Por este motivo, el funcionamiento de las RN no se adapta de la mejor manera a las necesidades de los datos y los resultados lo

demuestran.

La RL se caracteriza por ser un modelo bastante sencillo aunque, debido a la naturaleza de las señales disponibles, cumple con las expectativas y obtiene resultados muy positivos. Se evita la extracción de características, simplificando el algoritmo, ya que se trabaja directamente con el valor absoluto de la señal de entrada. Sin embargo, a través de las SVM podemos modificar algunos factores que, aunque complican un poco mas el modelo, nos permiten obtener resultados óptimos. Se representan los datos a partir de otro más relevantes y se puede modificar la ventana sobre la que se extraen dichas características. Esto permite la correcta adaptación a la forma de la señal. Finalmente se concluye que el modelo a partir de SVM da los resultados más satisfactorios para la clasificación automática que se buscaba.

6. Conclusions

In this chapter there will be a general discussion about the results obtained in this work and future work lines will be commented as well.

6.1. General conclusions

The goal of this work was to create an algorithm capable of automatically detect epileptic seizures. For such purpose, three approaches were discussed: Logistic Regression, Neural Networks and Support Vector Machines. Once all the models have been successfully developed, it can be concluded that the classification achieved by analysing the sensor data signals is remarkably high for all tested classifiers. The comparison were based on the analysis on certain performance measures such as the sensitivity or specificity. All the recorded crisis has been correctly classified but certain samples still remain undetected due to the classifiers performance.

The seizure-detecting device was user friendly. All patients were able to handle the device with no difficulties. Detection accuracy shows the feasibility of the system for practical implementation allowing doctors to have enough supporting information for treatment decisions, though there was some postures and movements that there should be taken into account before starting the record.

Although there was several features that could be extracted from the signals, the selection of only those relevant to the algorithms has been highly important. Also, the implementation of a sliding window for the extraction has proved to be really useful to improve the classification performance. The selection of the correct window size determined the correct behaviour of the models.

All the algorithm have achieved very accurate results which suggests that this method could be viable for deployment. Despite this, the inner performance of the ANN still remains unknown so even it has great results, it doesn't give the same confidence as the other models proposed. RL and SVM have proved to work also really good, but the processing time required for RL was extremely high and downplays the good results obtained. For those reasons, SVM seems to be the most

suitable algorithm.

6.2. Future Work

Taking this thesis as the starting points, there are some improvements that could be done. Possible variations may be added to the design of the models to improve their performance.

- In terms of processing time and as it has been previously commented, the RL model was extremely slow so the adaptation of the code to another **programming platform** may result in improved processing time.
- In all the proposed models, samples are treated independently so even though seizures are correctly detected, the starting and finishing stages remain undetected. The **Hidden Markov Model** is proposed to fix this problem. This model exploits the decisions taken by the classifier in the past, so it is a sequential approach. The advancement of the samples is treated as if it was a chain, so every present decision depends on the previous ones. Another possible solution is the development of a **post-processing stage** where the independence of the samples is treated. The creation of an algorithm capable of joining samples detected as attacks with a very few time difference as just one seizure.
- In case of continuing with this research, it could be really interesting to increase the number of patients in order to have a larger and better labeled epilepsy **dataset**. This would enable better training and more conclusive testing of the performance of the classifiers. It is very interesting to see how the models could generalize to patients with different seizures and to conclude which model is best to train individually each patient.
- The **environment** under study is another alternative that could be modified. In this thesis it has been used data recorded from patients during sleep hours, but the recording may be extended to all day so the detection of seizures would become a multiclass classification problem.

A. Planificación y Presupuesto

En este anexo se mostrará la planificación seguida durante el desarrollo del proyecto a través de un diagrama GANTT. Finalmente se muestra un resumen del presupuesto obtenido.

A.1. Planificación del Proyecto

Para la correcta realización de este trabajo se ha decidido dividirlo en varias fases. La duración de cada actividad se desglosa en la siguiente Tabla:

Fase	Tarea	Duración
Documentación	Estudio de la funcionalidad del sensor	10 horas
	Estudio de la enfermedad	10 horas
	Instalación de las herramientas de trabajo	5 horas
Fase Inicial	Comprobación del etiquetado de los videos	10 hours
	Creción de la Base de Datos	10 horas
	Familiarización con el tipo de señales recolectadas	10 horas
	Investigación sobre modelos de clasificación	20 horas
	Realización de pruebas sencillas	10 hours
Desarrollo del programa	Calibración de los sensores	10 horas
	Preprocesado	30 horas
	Implementación de los algoritmos	40 horas
Pruebas	Pruebas de entrenamiento y test	20 horas
	Optimización de los algoritmos	15 hours
Memoria	Familiarización con la herramienta LATEX	5 horas
	Redacción de la memoria	70 horas
	Correcciones	10 horas
	Preparación de la presentación oral	15 horas
TOTAL		300 horas

Tabla 7: Planificación del proyecto.

A.2. Diagrama de Gantt

Para ilustrar el orden cronológico seguido durante la realización del proyecto se utiliza el diagrama de Gantt de la Figura A.1.

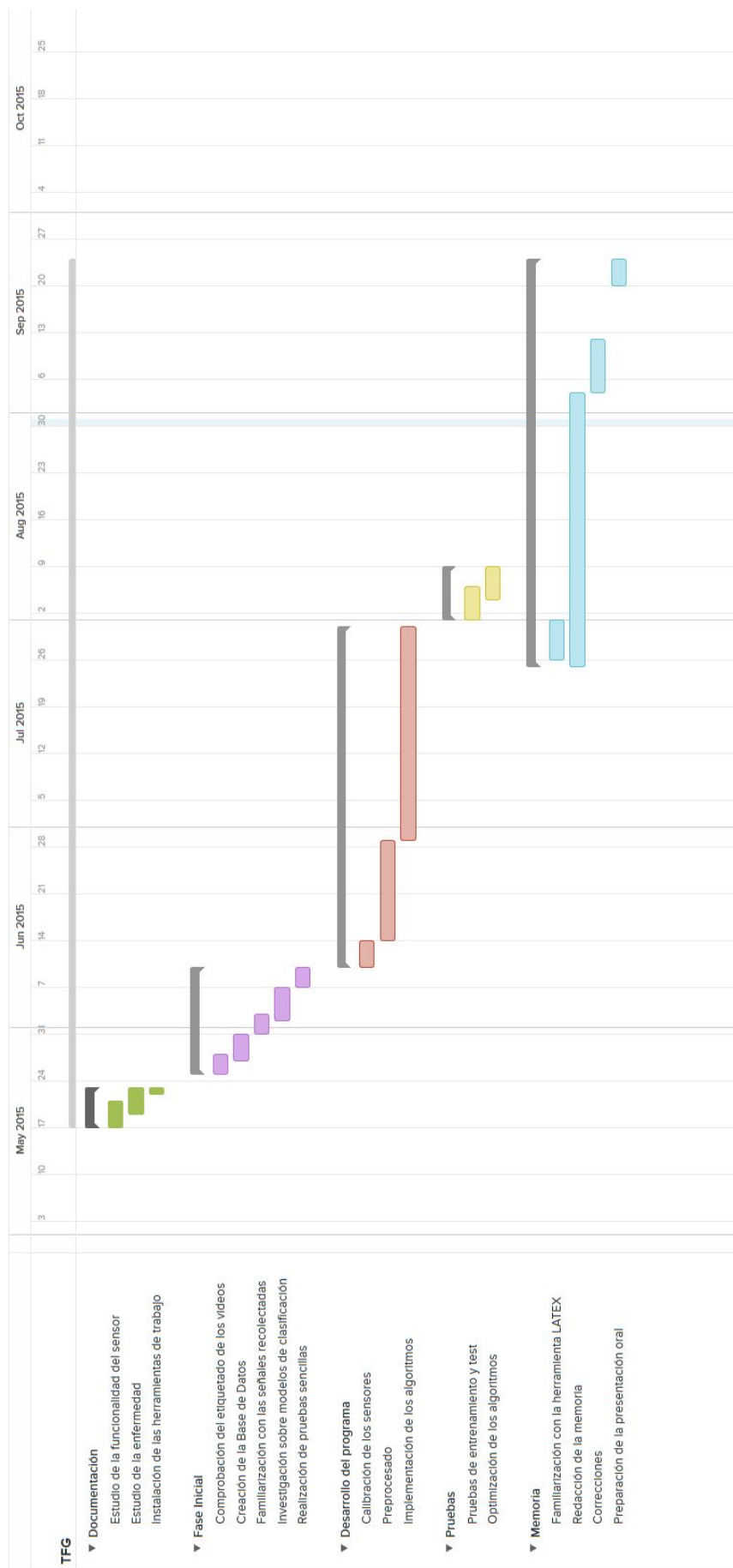


Figura A.1: Diagrama Gantt del proyecto.

A.3. Presupuesto

Los costes necesarios para la realización de este trabajo se desglosarán entre costes materiales y costes de personal.

A.3.1. Costes Materiales

El material utilizado para la realización de este proyecto incluye: un pack de sensores, dos prototipos de monitorización y un ordenador personal con sistema operativo Windows 7. La herramienta de programación ha sido MATLAB, con sus correspondientes gastos de licencia. El coste de proyecto mostrado en la Tabla 8 tiene en cuenta un periodo de amortización de 2 años para los aparatos tecnológicos y de un año para la licencia de MATLAB. La duración del proyecto ha sido de 4 meses:

Concepto	Precio (€) Total	Cantidad	Coste (€) Total	Coste (€) Proyecto
Prototipo de Monitorización	15000	2	30000	5000
Sensores Inerciales	359	4	1436	240
Ordenador	700	1	700	116
Licencia MATLAB	69	1	69	23
TOTAL				5379€

Tabla 8: Costes Materiales

A.3.2. Costes de Personal

La realización de este trabajo ha requerido la presencia de un jefe de proyecto y un ingeniero.

Concepto	Horas	Precio por Hora	Importe (€)
Jefe de Proyecto	20	90	1800
Ingeniero	280	60	16800
TOTAL	300		18600€

Tabla 9: Costes de Personal

A.3.3. Costes Totales

El coste total del proyecto se desglosa en la Tabla 10.

Concepto	Precio (€)
Costes Materiales	5379
Costes de Personal	18600
Costes Indirectos(20 %)	3720
<i>Subtotal</i>	<i>27700</i>
IVA (21 %)	5540
TOTAL	33240 €

Tabla 10: Costes Totales

El coste total del proyecto es de TREINTA Y TRES MIL DOS-CIENTOS CUARENTA EUROS.

Leganés, a 15 de Octubre de 2015

B. Summary: Epileptic Seizure Detection Through Inertial Sensors

This annex summarizes the bachelor thesis **Epileptic Seizure Detection Through Inertial Sensors** written by *Irene Tejado García* and supervised by *Victor Elvira Arregui*.

B.1. Introduction

Epilepsy is cataloged as one of the most common neurological disorders and patients often require long-term care which cost Europe almost €800 billion a year. From the technological point of view, the quick advancement in mobile technologies has brought the opportunity to acquire real-time data from patients spending large periods of time hospitalized in order to find a way to create a precise method to monitoring them from home.

The purpose of this thesis is the development of an algorithm capable of detecting epileptic seizures based on the data extracted from a wearable sensor placed on the patients under experiment. The data available has been provided by the Epilepsy Unit of the "Fundación Jiménez Díaz", lead by Dr. José Serrasota and contains records from nocturnal hours. Personal and medical records are protected by "La Ley Orgánica de Protección de Datos", in which a set of mandatory principles are considered regarding the patients privacy.

The seizure detector incorporates a sequence of steps that, applied to the data, facilitate the development of the algorithm and improve the results. The working method, depicted in Figure B.1, is divided into several parts that will be held for both training and testing set.

B.1.1. Shimmer Sensor

The selected sensor is portable and enables the capture, transmission and processing of body data detected in real time. It contains the tools to carry out the whole process. It must be lightweight so it does not compromise the normal development of the study.

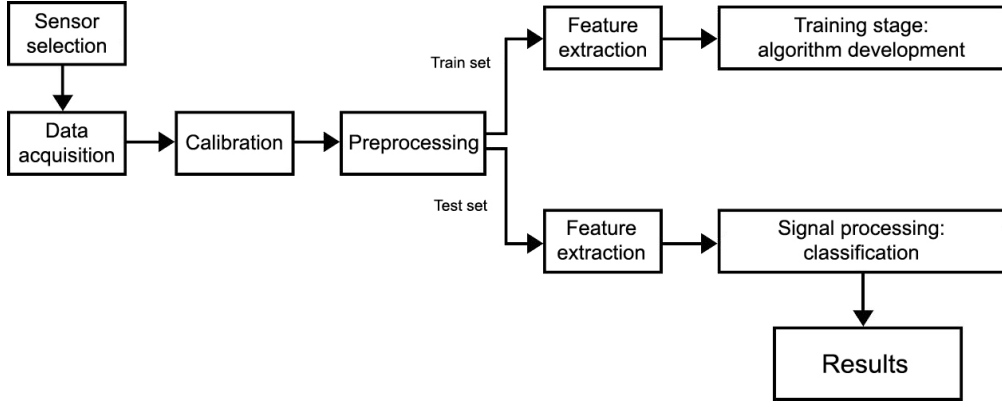


Figura B.1: Methodology Flowchart

Inertial Measurement Unit

The wireless sensor contains an electronic device called IMU (Inertial Measurement Unit) which measures the speed and rotation using an accelerometer, gyroscope and magnetometer. In each of the modules, records are obtained in the inertial axes X, Y and Z allowing a complete solution of the body data.

Getting the Inertial Signal

Inertial measurement units might be harmed by errors such as noise or misalignment. The captures are limited by values within a finite range while the physical phenomena observed is not.

For these reasons, it is necessary to define a rotation matrix that relates the sensitivity values obtained from the different inertial axes with the actual ones. Said matrix calculation is performed during the calibration process and it is defined as the matrix R .

For a tri-axial sensor and the phenomena to be detected \underline{v} , the final capture of the data is done as follows:

$$\underline{Y} = KR\underline{v} + \underline{b} + \underline{n} \quad (29)$$

where \underline{Y} is the sensor output, \underline{v} is the phenomena to be detected, KR are the matrixes for the scaling and rotating of the axes, \underline{b} is the bias vector and \underline{n} is the noise vector.

B.1.2. Data Collection

For data collection it has been designed a monitoring prototype in which two parts can be distinguish:

- **Recorder case:** it contains every device needed for taking data as depicted in Figure B.2: a computer display, wireless keyboard, camera, buttons, sensor base.
- **Computer:** there is a computer whose function is to run the application for conducting and display the processing information recorded by the case.



Figura B.2: Recorder Case

B.1.3. Data Base

The organization of the patients is done by implementing a database. Due to the Ley Orgánica 15/1999 de Protección de Datos, patients information can not be exposed so they are assigned an identification and registration number for every record they own. An example is shown in Table 11.

ID	Reg	File name	A01	A02	Date	Start time Record	Crisis time case
10	2	ID10_R2_00-000_ 20130920_050612-654	1	1	20-09-13	5:06:12	5:38:40 6:02:40 7:04:55

Tabla 11: Data Base

B.1.4. Calibration

Once all the patients data has been organized, the first step performed on the raw data is the calibration. This process is applied to

all the components of the sensor as they may be affected by different types of errors. When the sensor data is acquired, units are not suitable for processing. Calibration allows to get standard values and a calibrated data.

B.1.5. Preprocessing

The main purpose of preprocessing is to manipulate and transform the raw data in a data set more easily accessible. This stage includes all those analysis techniques to improve the quality of the data. It is responsible for segmenting and eliminating redundant information so machine learning models can be applied obtaining a better precision.

Outliers Removal

Outliers are defined as observations numerically far from the rest of data. Generally these values are related to errors or misleading information and therefore, they are excluded from the data set.

Through the signals obtained from the sensor, it follows that outliers are but peaks with minimum duration achieving much higher energy values. Applying a threshold, these disproportionate peaks are identified and replaced with an average value of the signal in positions adjacent to the outlier.

Low pass Filter

There also exist some peaks with a very low energy level that do not provide any information in order to extract features for pattern classification. Next step aims to eliminate those peaks and for such purpose, a low pass filter is used. By defining a cut-off frequency and applying a Butterworth filter, all signal components whose frequency is above the cutting will be attenuated and so it is achieved the elimination of level noise.

B.1.6. Features Extraction

Certain features are extracted from the data in order to represent it through other most significant values that help with the task of differentiating between seizures and steady.

The selection of those that maximize the help to get a correct response from the classifier is very important:

- **Mean:** the average of a set of values is calculated. This parameter is widely used for distributions probability.
- **Standard deviation:** it gives a measure of how data is dispersed from the average value. It can also be interpreted as a measure of uncertainty.
- **Maxima:** to facilitate the peaks detections, the maximum value of each subset of data is obtained.
- **Energy:** getting the energy value from the signals allows to set thresholds to detect crisis. The intensity of the movements is represented more clearly making the data interpretation easier.

The procedure to be followed is based on the creation of a **sliding window** that moves along the entire signal with an overlap of 50 % facilitating the characteristics extraction.

B.2. Signal Processing

The evaluation of the physical activity obtained through the sensors is possible thanks to the use of classifiers. Describing some phenomena by statistics, their goal is to assign a specific category to an input. More specifically, there will be special emphasis on **supervised learning**, in which the training set is labeled so the classifier knows beforehand to what type of category belongs each of the samples used during the training phase.

Proposed Classifiers

The choice of the classifier is one of the most important aspects to consider. Next, it will be explained the theory concerning the classifiers used in this project.

B.2.1. Logistic Regression

It is one of the greatest statistical tools for data analysis in clinical research. This model is used to predict the outcome of a dependent variable according to one or several explanatory variables. It is modeled through the logistic function shown in Figure B.3, which represents the relation between the regression model and the detection probability.

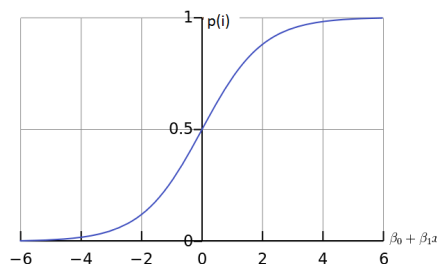


Figure B.3: Logistic Function

Its mathematical expression is as follows:

$$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)} \quad (30)$$

The development of this method has been done based on two approaches:

Data set	Regression Model
ACC	$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1)}$
ACC + GYR	$p_i = \frac{1}{1 + \exp -(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$

The result obtained is treated as if it was a hypothesis where a value close to "0" indicates a low probability of attack and a value close to "1" means high seizure probability. Finally, a threshold should be defined to select or discard those attacks in the data set.

B.2.2. Neural Networks

Neural Networks are a technique of automatic processing inspired by the way the nervous system works and, as such, tries to reproduce the behaviour of the brain. It is defined as a set of processing units called neurons, capable of storing information and connecting among

them.

Architecture

Each neuron receives a number of inputs p that are transmitted through a connection that multiplies them by an specific weight w , that indicates the neuron to which they are intended. To this product is added the bias value. Neural networks usually contain more than one layer to optimize the system as shown in Figure B.4.

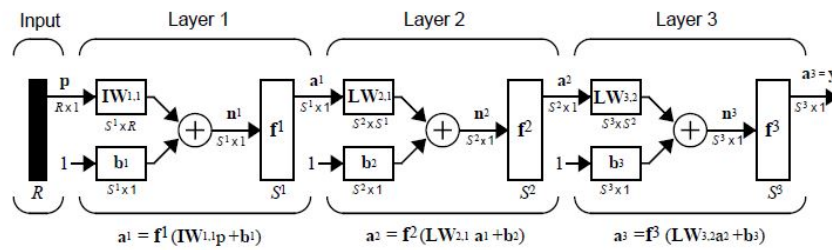


Figure B.4: Multiple layer Neural Network.

Applications

When the training samples and their corresponding labels are available, what is sought is to find a function that relates them to deduct labels from samples. However, this function is unknown and hypothesis appear instead. The goal of Neural Networks is to minimize the risk of those assumptions.

Advantages

Artificial Neural Networks are widely used in the medicine field. They have an adaptive learning system as a result of the training stage and are able to create its own data structure. They also have high error tolerance since the information that is distributed along all neurons has some redundancy. Finally, this type of machine learning system computes in parallel thus shortening the operation in real time and at low cost.

B.2.3. Support Vector Machines

The Support Vector Machines are a set of algorithms based on supervised learning and were developed to solve classification problems or pattern recognition.

SVMs theory focuses on mapping the entry points to a larger feature space in order to find an optimal hyperplane that separates them and maximize the margin between the different categories to classify. The vector formed by the closest points to the hyperplane is known as support vector.

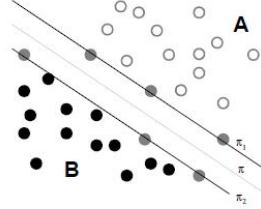


Figura B.5: Set of data and hyperplane

To be able to separate the data sets, it is necessary the use of a kernel function. In addition, it uses a parameter C that determines the exchange between the two goals of the model: maximize the margin and minimize the training errors.

Advantages and Disadvantages

The development of SVMs has very few parameters to be adjusted since it only depends on those data with more information. The kernel function contributes to the algorithm in a great generalization capability even if the training set is small.

However, when the complexity of the classifier increases, the system becomes very sensitive to small variations in the input values and could suffer *overfitting*. Another problem is the *overtraining*. This happens when the training data has been learnt such well that the classifier is unable to properly classify new data.

B.2.4. Performance Measures

There will be some concepts proposed to measure the performance of the mentioned models. where TP is the number of true positives,

Sensitivity (S) = $\frac{TP}{TP+FN}$	Especificity (E) = $\frac{TN}{TN+FP}$
Precision (P) = $\frac{TP}{TP+FP}$	F Measure = $2 * \frac{P*S}{P+S}$

FP is the number of false positives and TN and FN true negatives and false negatives respectively. Ideally, it is intended that the results

always have high sensitivity and specificity values. To evaluate the effectiveness of the system, both values are related through ROC curves as the one depicted in Figure B.6. The area under the curve will determine the effectiveness of the proposed model.

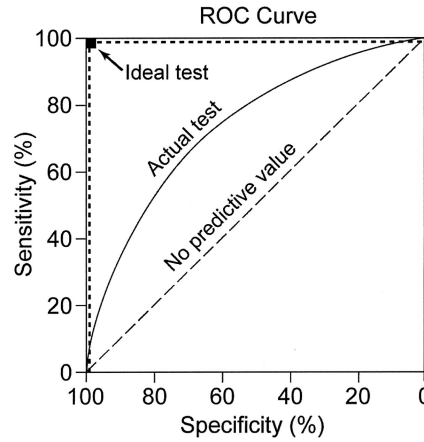


Figura B.6: ROC curve

B.3. Results

The results obtained are summarized in Table 12:

Model	Implementation	Processing Time	P	E	S	F
RL	ACC	High	1	1	0.952	0.975
	ACC+GYR	Very high	0.992	0.99	0.965	0.979
ANN	window = 20	Medium	0.826	0.999	0.690	0.752
	window = 100	Medium	0.951	1	0.963	0.956
SVM	window = 20	Low	0.998	0.999	0.989	0.993
	window = 100	Low	1	1	0.753	0.852

Tabla 12: Results Summary

It has been found that the size of the sliding window is a determining factor in the effectiveness of the algorithm. Also, since all the models proposed treat the samples independently, seizures are detected at its highest level. The beginning of the crisis is sometimes missed since this algorithms has a lack of memory to deduce an attack is coming.

Logistic Regression seems to have higher success than the other models proposed although a large number of samples were classified as false negative with this model. This category is not interesting in

terms of diagnosis since ignoring that the patient is having a seizure would be more damaging than obtaining false positive, where there is an alert that the patient is having a seizure but it is not real. Neural Networks reach a very precise classification but they vary widely from one training to another since the set division is random. To realize this project, there is actual data collected from patients who were admitted to monitor them for research purposes. The data set contains few seizures so the choice of the training set is extremely important and determines the effectiveness of the algorithm. Therefore, the operation of the neural networks is not adapted in the best way

Logistic Regression is characterized by a rather simple model and due to the nature of the signals available, gets very positive results. Also, avoids features extraction, simplifying the algorithm. However, through Support Vector Machine, we can achieve a more complex model allowing to optimize the results. Data is expressed from other more relevant values and the modification of the extraction window allow the correct adaptation to the shape of the signal. Finally, it is concluded that the model based on Support Vector Machine gives the most satisfactory results for the classification sought.

B.4. Conclusions and Future Work

The classification achieved by analysing the sensor data signals is remarkably high for all tested classifiers. The device was user friendly which shows the feasibility of the system for practical implementation allowing doctors to have enough supporting information for treatment decisions.

The selection of only those features relevant to the algorithm has been highly important and so has been the implementation of a sliding window.

Some variations may be added to the design of the classifiers to improve their performance:

- In terms of processing time, adapting the code to another programming platform may result in an improvement.
- In all the proposed models, samples are treated independently so the Hidden Markov Model is presented to fix this problem. This

model exploit the decisions taken by the classifier in the past. Also, the development of a post-processing stage would be suitable to treat the independence of the samples just creating an algorithm capable for joining samples detected as attacks with a very few time difference as just one seizure.

- The increase of the number of patients would enable better training and more conclusive testing on the performance. Also the environment under study may be extended.

Referencias

- [1] J. Christensen, M. Vestergaard, M. G. Pedersen, C. B. Pedersen, J. Olsen, and P. Sidenius, “Incidence and prevalence of epilepsy in denmark,” *Epilepsy research*, vol. 76, no. 1, pp. 60–65, 2007.
- [2] L. Orgánica, “Ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal,” *BOE de [Internet]*, vol. 14, 1999.
- [3] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, “Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe),” *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.
- [4] K. Cuppens, P. Karsmakers, A. Van de Vel, B. Bonroy, M. Milosevic, S. Luca, T. Croonenborghs, B. Ceulemans, L. Lagae, S. Van Huffel *et al.*, “Accelerometry-based home monitoring for detection of nocturnal hypermotor seizures based on novelty detection,” *Biomedical and Health Informatics, IEEE Journal of*, vol. 18, no. 3, pp. 1026–1033, 2014.
- [5] E. Ben-Menachem, “Medical management of refractory epilepsy—practical treatment with novel antiepileptic drugs,” *Epilepsia*, vol. 55, no. s1, pp. 3–8, 2014.
- [6] R. Ramani, “Vagus nerve stimulation therapy for seizures,” *Journal of neurosurgical anesthesiology*, vol. 20, no. 1, pp. 29–35, 2008.
- [7] M. Atkinson, K. Hari, K. Schaefer, and A. Shah, “Improving safety outcomes in the epilepsy monitoring unit,” *Seizure*, vol. 21, no. 2, pp. 124–127, 2012.
- [8] K. Devarajan, E. Jyostna, K. Jayasri, and V. Balasampath, “Eeg-based epilepsy detection and prediction,” *International Journal of Engineering and Technology*, vol. 6, no. 3, p. 212, 2014.
- [9] V. Villanueva, A. Gutierrez, M. Garcia, A. Beltrán, J. Palau, R. Conde, P. Smeyers, P. Rubio, E. Gómez, T. Rubio *et al.*, “Usefulness of video-eeg monitoring in patients with drugresistant

- epilepsy,” *Neurología (English Edition)*, vol. 26, no. 1, pp. 6–12, 2011.
- [10] A. Steventon, M. Bardsley, J. Billings, J. Dixon, H. Doll, S. Hirani, M. Cartwright, L. Rixon, M. Knapp, C. Henderson *et al.*, “Effect of telehealth on use of secondary care and mortality: findings from the whole system demonstrator cluster randomised trial,” *Bmj*, vol. 344, 2012.
 - [11] I. Conradsen, S. Beniczky, K. Hoppe, P. Wolf, and H. B. Sorensen, “Automated algorithm for generalized tonic-clonic epileptic seizure onset detection based on semg zero-crossing rate,” *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 2, pp. 579–585, 2012.
 - [12] L. Kumar and G. Srivastava, “Epilepsy detection using clustering techniques,” 2014.
 - [13] *IMU User Guide Revision 1.3*, Shimmer Realtime Technologies Ltd, 2014.
 - [14] L. Tan and J. Jiang, *Digital signal processing: fundamentals and applications*. Academic Press, 2013.
 - [15] R. A. Johnson, *Statistics: principles and methods*. John Wiley & Sons, 2009.
 - [16] C. C. Aggarwal, *Data classification: algorithms and applications*. CRC Press, 2014.
 - [17] A. M. S. Andrea Mannini, “Machine learning methods for classifying human physical activity from on-body accelerometers,” 2010.
 - [18] I. Schuster and P. Jähnichen, *Classification using Logistic Regression*. Universität Leipzig.
 - [19] W. G. Baxt, “Application of artificial neural networks to clinical medicine,” *The lancet*, vol. 346, no. 8983, pp. 1135–1138, 1995.
 - [20] M. Astion and P. Wilding, “The application of backpropagation neural networks to problems in pathology and laboratory medicine.” *Archives of pathology & laboratory medicine*, vol. 116, no. 10, pp. 995–1001, 1992.

- [21] S. Marsland, *Machine learning: an algorithmic perspective*. CRC press, 2014.
- [22] J. Li, “An empirical comparison between svms and anns for speech recognition.”
- [23] A. Alkan, E. Koklukaya, and A. Subasi, “Automatic seizure detection in eeg using logistic regression and artificial neural network,” *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167–176, 2005.